Deliverable nr D5.2 – Evaluation Plan & planning

Work Package: WP5
Task: T5.2, T5.3
Date of delivery
Contractual: 31-12-2008
Actual: 13-02-2009
Code name
Version: 1.1
Type of deliverable: Report
Security: PU

Contributors
Authors (Partner): Thomas Heider (LiNK), Rani Pinchuk (SAS), Lynn De Proft (SAS), Marlies Bitter (OUNL), Jo Boon (OUNL)
Contact Person: Thomas Heider
WP/Task responsible: WP5
EC Project Officer: Mr. Christian Wilk

Abstract (for dissemination)
Deliverable 5.2 is an output from WP5, User Requirements and Evaluation. D5.2 has two major objectives:

Evaluation Methodology Development.
Planning of the Evaluation Activities.

With the evaluation of idSpace, we will measure the characteristics of the platform, in the attempt to compose a general idea about its attributes and features. While doing so, we will unveil the points of strength and weakness.

This document is intended to be read by the people who are involved in the evaluation.

Keywords List
evaluation, method, usability, usefulness, Delphi
Table of Contents

1. Executive Summary .............................................................................................................. 5

2. Introduction .......................................................................................................................... 6
   2.1. Purpose of the Document ................................................................................................. 6
   2.2. Scope .................................................................................................................................. 6
   2.3. Used Documents ............................................................................................................... 7
   2.4. Definitions and Conventions ............................................................................................ 7
   2.5. Structure of the document ................................................................................................. 7

3. Evaluation Methodology ....................................................................................................... 8
   3.1. Determining the parameters of the project ....................................................................... 8
       3.1.1. The phases .................................................................................................................... 8
       3.1.2. What is the focus of the evaluation? ........................................................................... 9
       3.1.3. Usability ..................................................................................................................... 10
       3.1.4. Usefulness ................................................................................................................ 11
       3.1.5. What has to be evaluated? .......................................................................................... 11
       3.1.6. Objectives, requirements and constraints of IdSpace ................................................. 11
       3.1.7. Users of the evaluation results .................................................................................... 17
       3.1.8. Who are the evaluation participants ......................................................................... 17
       3.1.9. Context ...................................................................................................................... 17
   3.2. Selecting an evaluation scenario ...................................................................................... 18
   3.3. Evaluation methods ......................................................................................................... 19
       3.3.1. Phase 1 ...................................................................................................................... 19
       3.3.2. Phase 2 ...................................................................................................................... 22

4. Evaluation Plan ..................................................................................................................... 25
   4.1. Phase 1 ........................................................................................................................... 25
       4.1.1. Context ...................................................................................................................... 25
       4.1.2. Scenarios ................................................................................................................... 26
       4.1.3. Procedure .................................................................................................................. 26
       4.1.4. Processing the results ............................................................................................... 26
       4.1.5. The idSpace e-Delphi study ..................................................................................... 26
   4.2. Phase 2 ........................................................................................................................... 28
       4.2.1. Context ...................................................................................................................... 28
       4.2.2. Scenarios ................................................................................................................... 29
       4.2.3. Procedure .................................................................................................................. 30
       4.2.4. Processing the results ............................................................................................... 31

5. Conclusion ............................................................................................................................ 35

6. References ............................................................................................................................. 36

7. Appendix ............................................................................................................................... 37
   7.1. Questionnaire SUS - System Usability Scale ................................................................. 34
   7.2. Usability Heuristics by Nielsen (Nielsen, 1994) ............................................................... 36
   7.3. NASA TLX Paper and Pencil Version ............................................................................. 40
   7.4. Usability Training - Course Syllabus ............................................................................... 41
D5.2 Evaluation plan & planning
Executive Summary

The main objective of the evaluation of idSpace is to analyze the impact and viability of the technology developed, showing the efficiency, appropriateness, meaningfulness and usefulness of the idSpace platform as tool for supporting innovation and creativity.

This document serves as an outline for the preparation and planning of the evaluation of the idSpace platform. It has two major parts: the description of the evaluation methodology and the planning of the evaluation activities.

As the development of the isSpace platform is following the proposed cycles of the standard ISO 13407, the evaluation will also take place in different phases. The different phases are reflected in the document.

Evaluation Methodology Development: The evaluation methodology needs to ensure that the evaluation of technical functionality takes place in parallel with the evaluation of pedagogic soundness. Different approaches will be analyzed and combined into a framework that meets the specific requirements of the idSpace context. Constraints, success factors, quality factors and assessment criteria for the idSpace environment will be described and are the basis of the decision for the final evaluation methodology. The methodology includes 1. Base claims: Evaluation of ease of use (usability); 2. Distinguishing claims: Evaluation of idSpace as a tool that is facilitating more creative and innovative design processes.

Plan the Evaluation Activities: For each phase of the development process of the idSpace platform (planning, development and demonstration) concrete evaluation activities and tasks will be performed regarding the Evaluation Methodology. Different methods of acquiring usability feedback (e.g. user and expert walk-throughs) and application feedback data (e.g. effects on creativity) will be defined in details. The evaluation plan will be based on the ISO 9241-11 guidance on usability and will include guidelines for gathering user evaluation feedback applied by all partners performing evaluation activities.
Introduction

This document describes the planned evaluation of the idSpace platform and its underlying ideas. The impact and viability of the technology will be tested. The evaluation will provide evidence for the efficiency, appropriateness, meaningfulness and usefulness of the idSpace platform as a tool for supporting innovation and creativity. The bottom line here is that (continued) use of the idSpace platform should result in the generation of more ideas at a faster pace.

This evaluation is focused on two groups of claims that idSpace wants to satisfy:

- **Base claims of the platform** - Do the technology and design support innovative and creative processes?
- **Distinguishing claims** - Do the technology and design enhance creative processes, i.e. do groups working with it generate more and better ideas more quickly?

This document will therefore provide:

- An overview of the planned evaluation of the idSpace platform, on the basis of tests carried out by users and experts.

*Purpose of the Document*

Deliverable 5.2 is an output from WP5, User Requirements and Evaluation. D5.2 has two major objectives:

- Evaluation Methodology Development.
- Planning of the Evaluation Activities.

With the evaluation of idSpace, we will measure the characteristics of the platform, in the attempt to compose a general idea about its attributes and features. While doing so, we will unveil the points of strength and weakness.

This document is intended to be read by the people who are involved in the evaluation.

*Scope*

This document (D5.2.) contains two tasks from WP5:

- T5.2 - Evaluation Methodology Development;
- T5.3 - Plan the Evaluation Activities.

D5.2 will be the basis for the tasks:

---

1. With underlying ideas we mean the new approaches of idSpace (developed in the research work packages 1-3), e.g., the combination of different creativity techniques, the conjunction of learning and creativity techniques, and the application of context awareness. The performance of underlying ideas will be tested implicitly.

2. idSpace in this document always refers to the web-based idSpace platform.
D5.2 Evaluation plan & planning

- T5.4 - First evaluation;
- T5.5 - Second evaluation.

The results of the evaluation will be published in documents:
- D5.3 - Report of evaluation results, v1;
- D5.4 - Report of evaluation results, v2.

**Used Documents**

<table>
<thead>
<tr>
<th>Document</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO 9241-11</td>
<td>Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability</td>
</tr>
<tr>
<td>ISO 13407</td>
<td>Human centred design processes for interactive systems</td>
</tr>
<tr>
<td>ELPUB 105 10107</td>
<td>Handbook for practical usability engineering in IE projects. (Melchior et al. 19995)</td>
</tr>
</tbody>
</table>

**Definitions and Conventions**

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue Sky Thinking</td>
<td>Open minded thinking, which is not constraint by any borders.</td>
</tr>
<tr>
<td>Context of use</td>
<td>Users, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used.</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>Accuracy and completeness with which users achieve specified goals.</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Resources expended in relation to the accuracy and completeness with which users achieve goals.</td>
</tr>
<tr>
<td>Goal</td>
<td>Intended outcome.</td>
</tr>
<tr>
<td>ITT</td>
<td>Invitation To Tender</td>
</tr>
<tr>
<td>Measure</td>
<td>Value resulting from measurement and the process used to obtain that</td>
</tr>
<tr>
<td>Product</td>
<td>Part of the equipment (hardware, software and materials) for which usability is to be specified or evaluated.</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Freedom from discomfort, and positive attitudes towards the use of the product.</td>
</tr>
<tr>
<td>Task</td>
<td>Activities required to achieve a goal.</td>
</tr>
<tr>
<td>UI</td>
<td>User Interface.</td>
</tr>
<tr>
<td>Usability</td>
<td>Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. NOTE- See Annex D for other approaches to usability.</td>
</tr>
<tr>
<td>User</td>
<td>Person who interacts with the product.</td>
</tr>
<tr>
<td>Work system</td>
<td>System, consisting of users, equipment, tasks and a physical and social environment, for the purpose of achieving particular goals. Note that the context of use consists of those components of the work system which are treated as given when specifying or measuring usability.</td>
</tr>
</tbody>
</table>
Structure of the document

Chapter 3 contains the description of the evaluation methodology. Section 3.1. will present the parameter of the project and the focus of the evaluation. Furthermore it will show the phases of the idSpace development and how the evaluation will be integrated in the development process. In addition it will be identified what we will measure in the evaluation. Section 3.2 defines the constraints over the evaluation scenarios and Section 3.3 examines possible evaluation methods for the different phases and justifies our selection. In Chapter 4 the evaluation plan for both phases is illustrated.

Evaluation Methodology

This chapter discusses the methodologies used in order to evaluate the idSpace platform. In order to understand exactly what is being measured, the parameters of the project are determined. Then, the way the evaluation scenarios are chosen is described. Finally, different existing methodologies are examined, and the ones to be used are chosen.

Determining the parameters of the project

The phases

The development of the idSpace platform is carried out in 3 phases. First, an existing web-based collaboration platform (Microcosmos) and an existing tool for the representation and interchange of knowledge (Topic Maps, Ontopia) are integrated. This constitutes the first development effort in the project, resulting in the first prototype. No explicit design document underlies this prototype as it is an integration of existing products. The challenge here is to integrate them so that further development is feasible. While the prototype is being built, theoretical insights on adequate pedagogies, on idea generation and transformation and the usage of context are investigated and formulated as design constraints.

Together with user requirements resulting from collected use cases and from experience with the first prototype, the design constraints constitute the first design document. This document drives the development of the second prototype. This prototype is again evaluated, the results of which lead to a revision of the original design document and, consequently, a revision of the existing prototype, resulting in the third prototype. This prototype will be evaluated cursorily only, which will lead to an amplification of the second evaluation report.

The idSpace project is - with respect to the implementation of the platform - structured in three main phases. Therefore, we have also three different evaluation tasks.

<table>
<thead>
<tr>
<th>Phases</th>
<th>Time period</th>
<th>Evaluation tasks</th>
<th>Time period</th>
</tr>
</thead>
</table>
User Requirements definition.

Prototype version 2:
Integration of the technologies of WP 1, 2, 3.
Incorporation of the user requirements of D5.1.
Adjustment of the platform, based on the evaluation of v1.

<table>
<thead>
<tr>
<th>M14 – M18</th>
<th>Evaluation of Usability and Usefulness of prototype v2.</th>
<th>M19 - M20</th>
</tr>
</thead>
</table>

Prototype version 3:
Final adjustment of the platform, based on the evaluation of v2.

<table>
<thead>
<tr>
<th>M20 - M24</th>
<th>The impact of the platform will be investigated by WP6 in its exploitation and dissemination plan.</th>
<th>M24++</th>
</tr>
</thead>
</table>

The evaluation of the different phases of the project will have different foci:

**Prototype v1:**
- Focus on problem discovery
  - More formal testing

**Prototype v2:**
- Focus on measurement
  - Comparison against objectives
  - Comparison of products
  - More formal testing

**What is the focus of the evaluation?**

Evaluation can focus on different aspects:

- **Information content**
  The user evaluation of the information content is performed to assess the correctness of the information as well as the adequateness of the media (audio, video, graphic, text) selected for the presentation of the information.

- **Functionality of the user interface**
  The user evaluation of the user interface looks at features and properties of the dialogue and presentation components between the user and the user-interface of the platform. Assessed is the quality of use of the functions provided for browsing and navigating through the information content, for searching, selecting, retrieving and storing of information, and other functions.

- **Load imposed on the user**
  The assessment of cognitive load imposed on the user looks at the properties of the tasks which the user intends to perform with the platform and the mental effort these impose on the user. These are the properties of information content and the functionality of the application and the nature of
problem solving and decision making the user performs on the basis of this information.

The cognitive load theory (CLT, defined by Sweller) (Sweller, Van Merriënboer, & Paas, 1998 and Clark, Nguyen & Sweller, 2006) aims at providing guidelines and instructions to optimize learning. It builds on human information processing characteristics and suggests that one needs to recognize inherent limitations of concurrent working memory load on learning. For the idSpace platform, measuring perceived mental effort is relevant at both evaluation moments (the 1st and 2nd phase referred to in this document). Applying empirically-based guidelines from the CLT theory aims at the decrease of extraneous cognitive load, i.e., the working memory load experienced by learners as they interact with (instructional/learning) material during learning. Pedagogical recommendations inspired by CLT create affording circumstances for the knowledge processes involved in collaborative creativity using the idSpace environment. It helps to focus idSpace user's attention to connect to cognitive schemata (the so called germane cognitive load). Measuring the actual mental effort that idSpace users experience are needed to observe its effects

- **Stress perceived by the user when using the platform**
  Stress in terms of cognitive load and subjectively experienced discomfort is a function of the load on the individual user, and factors such as individual differences, learning state, and fatigue.

- **Benefits of the platform**
  The user evaluation of the platform is performed to gain insight in its benefits. This encompasses the way creativity and collaborative learning are enhanced. Quantitatively this can be measured by data on the number of interventions per participant, the recommendations used, and the number of interventions about the same topic, or the number of different topics in a given time interval. Qualitative data will encompass the attractiveness of the platform and the extent to which it challenges participants to express ideas, the perceived value of learning, and collaboration support.

**Focus of Phase 1**
- Functionality of the user interface

**Focus of Phase 2**
- Information content
- Load impose on the user
- Stress perceived by the user when using the platform
- Benefits of the platform

**Usability**

The Usability assessment plan describes the activities to assess the extent to which the IdSpace system can be used to achieve specified goals with effectiveness, efficiency and satisfaction.
Usability definition

One of the key questions when assessing usability and determining the objectives of usability assessment is the definition of the ‘usability’ concept. Many definitions of usability exist. These different definitions have some features in common. They all have the attributes: effectiveness, efficiency and satisfaction.

These attributes are part of the ISO 9241-11 definition (Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability) which will be used further on to assess the usability of the IdSpace system.

ISO 9241-11 defines usability and explains how to identify the information which is necessary to take into account when specifying or evaluating usability of a visual display terminal in terms of measures of user performance and satisfaction. The guidance includes procedures for measuring usability but does not detail all the activities to be undertaken. ISO 9241-11 applies to office work with visual display terminals. Design for usability will contribute positively to ergonomic objectives, such as the reduction of possible adverse effects of use on human health, safety and performance. ISO 9241-11 does not cover the processes of system development. Human-centred design processes for interactive systems are described in ISO 13407.

ISO 9241-11 defines usability as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

There are 3 different prototypes of the Idspace system planned. Usability assessment can be conducted at these 3 prototypes. We expect that usability assessment has the most important function at the first and second prototype. Besides usability assessment itself, development of the instrumentation and analyses of the results of the usability assessment have to be planned.

Usefulness

Definition of usefulness for the evaluation in the second phase.

Usefulness is concerned with the functionality of a system, while the usability is more about how easy the user can access the functions. Useful features enable users to reach their goals by providing the tools needed to perform specific tasks.

What has to be evaluated?

Subject of the evaluation is the IdSpace web-based platform and toolset that aims to support distributed collaboration in creativity, which is the process of generating new ideas, and innovation, which is the translation of the ideas into new products, services, or production methods. The evaluation framework will provide a methodology and the evaluation activity planning that will capture the success of all these elements.
The precise understanding of the parameters of the project is the prerequisite for planning the user evaluation process and for an agreement with users of the evaluation results. Therefore, in this first stage of the user evaluation process, the parameters have to be determined. The objectives, requirements, and constraints for the project must be determined and the users of the user evaluation results must be identified.

**Objectives, requirements and constraints of IdSpace**

The main claims of IdSpace that have to be proven by the evaluation are:

- The technology and design supports innovative and creative processes!
- The technology and design enhances creativity processes!

To feed the design of prototype v2, we analyzed use cases and elicited the corresponding user requirements (Deliverable D5.1 of WP5). The final list consists of 101 items, representing concrete User Requirements. The list is structured around the following categories:

- Preparation of the Ideation Session (Phase 1)
- Ideation Event / Creativity Session (Phase 2)
- Post processing of the Ideation Session (Phase 3)
- Accessibility
- Asynchronous/Synchronous Communication
- Collaboration and Content Sharing
- Content Dependent Functionality
- Creativity Techniques
- Miscellaneous Functionality
- Learning/Pedagogical Issues
- Personalization
- Tools
- Tracking
- Usability

Of course, it will not be possible to evaluate and proof every single user requirement. What has to be evaluated is the accomplishment of the prime objectives of IdSpace.

**Prime Objectives**

A major goal of the evaluation of prototype v2 is to proof the accomplishment of the prime objectives of IdSpace. For that reason the following section contains the enumerated objectives of the "prime objective session" in Brussels and the extracted features that have to be measured with the evaluation. How the features have to be measured will be described in Section 4, based on the identified methods of Section 3.3.
IdSpace prime objectives => features to be measured by evaluation

Web-based collaborative platform for distributed usage

The idSpace platform is a concurrent learning- and work environment. It's a web-based environment for collaboration on the invention of creative and innovative products. It includes an awareness model for the collaboration of teams that provides information on activities of persons ideas.

- **Usability Measurement**: distributed collaborative working
- **Usefulness Measurement**: distributed collaborative working

Capturing, organizing and reusing ideas

IdSpace keeps track of the ideas/contributions and the process. It stores the information in an easy accessible way. When creating a new project, it reacts on keywords with information from historical projects that are stored. The environment has tools to generate ideas, reuse them, take them apart, criticize them, or even reject them.

- **Usability Measurement**: generation, handling, and reusing of ideas
- **Usefulness Measurement**: comparing experts and non expert using the system

Innovative products design

The scope is the solving of ill-defined problems in the context of NPD that are new and cannot be solved with already known procedures.

- **Usability Measurement**: time to clarify the problem, time to solve problem
- **Usefulness Measurement**: quality of new ideas

Context aware support

IdSpace is an environment that is aware of the problems that the team is occupied with and which generates available knowledge and recommendations for tackling the problem.

You can find information about
- Tips and guidance on how to use IdSpace
- Several creative techniques, (what to use, when)
- Optimal team composition, based on actual task and phase
- Phases in a process and the actions

- **Usability Measurement**: ease of use and satisfaction with the platform
- **Usefulness Measurement**: number of ideas reused, appropriateness of ideas reused, i.e. integration of re-used ideas in solution

Learning (about the subject domain) (strategies)

IdSpace provides learner support. It supports informal learning and learning in function of the performance the team should deliver. It gives tips during the ideation and guidance through the product design process. It supports learning on-the-

---

3 including effective learning in function of work

idSpace - 2008-216199
job/learning by doing. For this purpose it gives among others recommendations on smart use of creativity techniques and guidelines for creativity sessions. IdSpace supports informal learning on the subject domain and formal learning on tools and techniques.

- **Usability Measurement:** cognitive ease of use and satisfaction with learning process and outcomes
- **Usefulness Measurement:** Number of recommendations used and satisfaction with them. Number of ideas reused from peers, from stored ideas which are integrated in the final proposition. Mental effort, cognitive load of learning

(To optimize learning idSpace recommendations will use principles of cognitive load theory. In the evaluation subjective methods (CL rating scales the CLT index) will be used to measure cognitive load on the user to measure mental effort required to use idSpace functionalities.)

**Topic maps**

IdSpace enables multiple representations and manipulations of ideas (free writing, imagination and visualization as well as collages like mood boards, enrichment of information via tags and using the concept of topic maps to support relations between the (various components) of ideas.

- **Usability Measurement:** handling of topic maps

**Creativity techniques**

IdSpace is primarily a work environment with support for main pedagogical strategies and creative techniques for individual and collective ideation.

- **Usability Measurement:** respective creativity rooms (techniques), satisfaction of using different creativity rooms
- **Usefulness Measurement:** quality and/or number of new ideas

**Multi-perspective (different perspectives on different techniques)**

IdSpace permits experimentation (supported to this end several techniques). It supports collaboration, communication, documentation, semi-automatic tagging and saving.

- **Usability Measurement:** number of interactions on ‘green’ ideas before a conclusion is reached
- **Usefulness Measurement:** number of perspectives taken into account before the conclusion is reached

**Domain aware users**

IdSpace provides a generic basis with functionalities to tailor the work environment to fit the purpose of creative ideation in dedicated setting: hence it is possible to adapt and compose the environment for the team and company.

- **Usability Measurement:** adaptable workspaces
Main requirements

A number of important features that have to be evaluated can be derived from the intended abstract use case of IdSpace. How these features can be evaluated (e.g. formative or summative) will be decided later in the document.

The Main Elements of the Innovation Process (from deliverable D5.1):

<table>
<thead>
<tr>
<th>Ideation Step</th>
<th>Evaluation Measurement (How well are these features provided?)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Preparation</td>
<td></td>
</tr>
<tr>
<td>• Problem definition / topic clarification</td>
<td>Problem definition</td>
</tr>
<tr>
<td>• Identifying / contacting / inviting the participants</td>
<td>Group composition</td>
</tr>
<tr>
<td>• Defining the objectives and the timeframe</td>
<td>Goal setting</td>
</tr>
<tr>
<td>• Ensure that everyone understands the topic and objective</td>
<td></td>
</tr>
<tr>
<td>• Provide context and additional material</td>
<td>Provision of Information</td>
</tr>
<tr>
<td>• Establish a coherent scope within the team</td>
<td></td>
</tr>
</tbody>
</table>
Table 1 - Abstract use case and the corresponding evaluation features

<table>
<thead>
<tr>
<th>Ideation Step</th>
<th>Evaluation Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2. Ideation Event / Creativity Session</strong></td>
<td>(How well are these features provided?)</td>
</tr>
<tr>
<td>• Choosing a useful systematic</td>
<td></td>
</tr>
<tr>
<td>• Using an innovation room</td>
<td>Innovation rooms</td>
</tr>
<tr>
<td>• Choosing an appropriate creativity technique</td>
<td>Appropriate creativity techniques</td>
</tr>
<tr>
<td>• Generating 'wild' ideas</td>
<td>Number and quality of ideas</td>
</tr>
<tr>
<td>• People must learn from each other</td>
<td>Understanding of ideas from the group</td>
</tr>
<tr>
<td>o Clarifying the underlying rationale for their ideas</td>
<td></td>
</tr>
<tr>
<td>o By contextualizing them</td>
<td></td>
</tr>
<tr>
<td>• Sharing ideas</td>
<td></td>
</tr>
<tr>
<td>• Constructively critique of ideas</td>
<td></td>
</tr>
<tr>
<td>• Critically assess of ideas</td>
<td>Assessment of Ideas (weighting of solutions)</td>
</tr>
<tr>
<td>o Their innovative value</td>
<td></td>
</tr>
<tr>
<td>o Look for design oddities</td>
<td></td>
</tr>
<tr>
<td>o Contradictions</td>
<td></td>
</tr>
<tr>
<td>• Logging</td>
<td></td>
</tr>
<tr>
<td><strong>3. Post processing</strong></td>
<td></td>
</tr>
<tr>
<td>• Compiling minutes</td>
<td>Compiling minutes</td>
</tr>
<tr>
<td>• Evaluating, ranking of generated ideas</td>
<td>Ranking of ideas</td>
</tr>
<tr>
<td>• Finalizing new ideas or designs</td>
<td></td>
</tr>
<tr>
<td>• Ideas need to be transformed to new products or product features.</td>
<td></td>
</tr>
<tr>
<td>• Alignment with broader goals (e.g. generating value for an organization)</td>
<td></td>
</tr>
<tr>
<td>• Goal achieved?</td>
<td></td>
</tr>
<tr>
<td>o No: Start iteration.</td>
<td></td>
</tr>
<tr>
<td>o Yes: Communicating to superior or customer</td>
<td></td>
</tr>
</tbody>
</table>
Users of the evaluation results

Different developers may need the evaluation results for different purposes. The results of the evaluation of the first prototype will mainly influence the design of the second prototype. Therefore, the developers of WP4 are the users of the first evaluation results. As the first evaluation is most of all about usability, it shall help the developers of WP4 to improve the ease of use of the platform.

The second prototype will also incorporate functionality originated from the theoretical perspectives of the WPs 1, 2 and 3.

Who are the evaluation participants

This section discusses the evaluation participants. Note that the reasons for the choices (the different constraints) are presented, but the choices themselves are presented in chapter 4.

As explained later (see section Phase 1 on page 21), in the first phase of the evaluation, no usability experience can be really measured, as the prototype is simply a result of an integration of two applications. Therefore, evaluation will be done by Usability Experts.

The Usability Experts should be able to assess the effectiveness and efficiency of a software product, by following established procedures using usability heuristics. In addition, the Usability Expert should be able to suggest improvements to the software UI.

The Usability Experts will also support the next phase as evaluation coordinators who prepare, manage and assess the results of the evaluation sessions. In that second phase, the prototype can be evaluated for both usability and usefulness. Different groups of potential users are envisaged:

- A development team will evaluate the tool on usefulness in solving real development problems in a creative way.
- A strategy team will evaluate the tool on usefulness in producing creative solutions in a blue sky thinking context.
- Groups of students will evaluate the tool on usability and usefulness by working on problems provided by the evaluation coordinators.

Depending on the evaluation settings, availability of the users, and defined scenarios, a mixture of user group representatives can also be considered to further diversify evaluation results.

Context

This section discusses different constraints related to the environment, the occasions when the tool is used and the platform of the software.
User groups
Electronic information services and products aim at different users. Mainly two categories of users can be distinguished [10]:
• users of technology who produce information and make it available
• their clients, in this report called the end users or just users, who are mainly from the general public and the business community.
Different usability engineering aspects must be assumed for these groups. Users of technology are involved in the production, dissemination, and retrieval of electronic information. Their role as a user of technology may include the role of being an end user as well. For example, a user of technology who is involved in the production of a multi-media electronic newspaper may be considered an end user who retrieves assets from an assets database while other users of technology are responsible for the creation, update and maintenance of this assets database.
The focus on different types of user groups is suggested because important features of electronic information services and products have an effect on evaluation and validation criteria and on cost / benefit assessments.

Tasks users intend to perform
Evaluation needs to take into account the tasks users intend to perform with the electronic information product or service.
Different groups of users may perform different tasks with one and the same application. Regarding only the functionality of the electronic information product or service, you can neither tell whether the functionality is sufficient or insufficient, nor can you detect that there is too much functionality which the user does not need. Based on task analysis results, the adequateness of the electronic information application for the user tasks can be estimated. Some users may use the application on a regular basis, others only occasionally. A best guess about the frequency of use and the execution of tasks with the electronic information application should be made for a defined time interval (total life time of the application, per day, per week, or other interval).

Context of use
Finally, evaluation must be performed with reference to the context in which the electronic information service or product will be used. The context of use is defined by:
• The environment in which an electronic information product or service is used: at home, in public places, at the workplace. Noisy environments and lighting conditions may impair the perceived quality of the application.
• The properties of the technology into which the electronic information application will be integrated.

Selecting an evaluation scenario
This section defines the constraints over the evaluation scenarios – it answers the question “what is a good scenario”.

While evaluating the prototype, and especially in the second evaluation phase, several evaluation scenarios have to be prepared. Those scenarios provide a basis and generate context for the subsequent usability testing.
The group of chosen evaluation scenarios has to include the following features (although not each scenario needs to include all the features):

- Require a creative solution.
- The scenarios should be realistic.
- Allow for controlled measurement and evaluation.
- A wide range of situations should be covered, not just the common ones or those of most interest to the design team.
- Problem situations should be included that test the system concept, not only straightforward scenarios.
- No reference to user interface features that should be used by the users.
- Allow for sufficient coverage to identify repeated usability issues.

Different evaluation scenarios will be developed that are representative for the various user groups considered, using the system to achieve a specific outcome under specified circumstances over a certain time interval.

**Evaluation methods**

In this section we examine the possible Evaluation Methods for the different phases.

**Phase 1**

**Nature**: In Phase 1, we evaluate Prototype V1 which is an early integration of applications. Following this integration, an iterative design process will commence. This means that the evaluation is the first input for this process. Because this is an early integration, no usability experience can be really measured. In addition, in this stage there will be no formal design documentation available.

**Scope of Evaluation**: The evaluation is done in order to improve the interface functionality and design and to identify errors (formative evaluation). In addition, because of this is an early design phase, usability inspection method (with expert evaluator) will be used rather than usability testing method (with end-user).

**Rationale**: Since an iterative design process is used, the emphasis will be on obtaining feedback (listing usability defects) which can be used to improve the design.

**Possible Usability Inspection Methods for Phase 1**

- **Heuristic Evaluation**
  Heuristic evaluation (Nielsen, 1994) is a usability engineering method to find the usability problems in a user interface design so that they can be attended to as part of an iterative design process. Heuristic evaluation involves having a small set of evaluators examine the interface and judge its compliance with recognized usability “heuristics”.

  - The evaluation is done independently – each expert evaluates the software by himself.
  - No quantitative data is collected.
  - Can be conducted remotely.
ISO9241-11 Usability Measures covered:

- Effectiveness: Yes
- Efficiency: Yes
- Satisfaction: No (but can be covered during the debriefing session)

**Cognitive Walkthrough**
The cognitive walkthrough method is a usability inspection method focusing on how easy it is for users to accomplish tasks with the system. This method involves one evaluator or a group of evaluators who inspect a user interface by going through a set of tasks and evaluate its understandability and ease of learning.

- This method is especially recommended for systems that are self-explanatory (where no explicit learning is needed).
- No quantitative data is collected.
- Cannot be conducted remotely.

ISO9241-11 Usability Measures covered:

- Effectiveness: Yes
- Efficiency: No
- Satisfaction: No

**Pluralistic Walkthrough**
The method centers around using a group of users, developers and usability professionals to step through a task scenario, discussing usability issues associated with dialog elements involved in the scenario steps.

- This method is especially recommended for systems that are self-explanatory (where no explicit learning is needed).
- No quantitative data is collected.
- Cannot be conducted remotely.

ISO9241-11 Usability Measures covered:

- Effectiveness: Yes
- Efficiency: No
- Satisfaction: Yes
**Recommended Method: Heuristic Evaluation**

- Quick feedback to designers.
- Suitable for early stage of development process.
- Observer assistance possible.
- Use of recognized list of general usability heuristics, and category specific heuristics.
- Defined output list.
- Emphasis on functionality and design.
- Usability measures: efficiency, effectiveness.
- Can be conducted remotely.
- 4 - 6 usability experts required, no users, no developers.

Heuristic evaluation is performed by having each individual evaluator inspect the interface alone. Only after all evaluations have been completed, the evaluators are allowed to communicate and have their findings aggregated. During the evaluation session, the evaluator goes through the interface several times and inspects the various dialogue elements and compares them with a list of recognized usability principles (heuristics). The output from using the heuristic evaluation method is a list of usability problems in the interface with references to those usability principles that were violated by the design in each case in the opinion of the evaluator.

**Drawbacks of the method:**

Heuristic evaluation does not provide a systematic way to generate fixes for the usability problems or a way to assess the probable quality of any redesigns. A possible solution is to conclude the evaluation with a debriefing session including the experts, the potential observer and the design team. The debriefing can focus on possible redesigns to address the major usability problems, general problematic aspects of the interface design, as well as positive and negative aspects of the design.

Another drawback is that there is no feedback on interaction between system and user. However, at this early stage, such feedback cannot be collected.

**Personnel involved:**

- 4-6 experts, individual evaluation sessions (Approx. 4 hours per session).
- Optional: 1 observer per evaluation.
- 0 Developers.
- 0 users.

**e-Delphi study**

Apart from the v1 and v2 evaluations, we plan to organize an e-Delphi (Bitter-Rijpkema et al., 2002, Dick, 2000, Linstone & Turoff, 2002) study. The aim of this study is to gather data on learning effects and usability in the workplace, complementary to the other evaluation methods proposed in this deliverable.

The added value of the Delphi study is that it surfaces open and original feedback from a representative expert audience on idSpace plans such as design assumptions and assumptions about the way idSpace will solve problems in daily real life practices.
Therefore we propose a Delphi study as an effective way to additionally capture relevant insight from representative practitioner’s audience of experts in the idSpace domain and representatives of the target user audience. It aims at surfacing ideas not yet written down in research reports. The Delphi method lends itself especially for the idSpace project focussing on creativity in new product design. It provides an open format to surface feedback from a representative expert population. Since the expert population has busy agenda’s and is spread across Europe we propose to organize an electronic Delphi study to enable experts to contribute in a flexible way not disturbing their full agenda’s. (You will find the details of the e-Delphi study in section 0.)

**Phase 2**

**Nature:** In Phase 2, we evaluate Prototype V2 which is the result of the iterative design process which was based on the inputs from the 1st Phase evaluation. In this phase the design documentation is available, as well as an early User Manual. The evaluation in this phase is an evaluation of the usefulness of the software.

**Scope of evaluation:** Asses if user and organizational objectives have been achieved – check that required practice has been followed (summative evaluation). The output should emphasize properties of the interaction between the user and the system.

**Rationale:** In this phase, the software should prove to be useful as a support tool for creativity, and therefore the usefulness of the software is in the center of this evaluation.

**Possible Usability Evaluation Methods for Phase 2:**

- **Heuristic Evaluation**
  See above in the section Phase 1 on page 21.

- **Think-aloud protocol**
  Think aloud protocols involve participants thinking aloud as they are performing a set of specified tasks with the system. The users are requested to vocalize his thoughts, feelings, and opinions while interacting with the system.

  - **Drawbacks:** The methodology contributes to a cognitive overload, which might affect the evaluation results.
  - The methodology introduce usage context which is not natural.

**ISO9241-11 Usability Measures covered:**

- **Effectiveness:** Yes
- **Efficiency:** No
- **Satisfaction:** Yes
Subjective Assessment (Questionnaire)

User evaluation instrument designed as a questionnaire such as Perceived Usefulness and Ease of Use (PUEU), Questionnaire for User Interface Satisfaction (QUIS), etc. It provides feedback from the user point of view.

- No quantitative data is collected

In the literature we can find various questionnaires for assessing the perceived usability of an interactive system. Tullis and Stetson (Tullis & Stetson, 2004) compared 5 different questionnaires for assessing the usability of websites. Three of the investigated questionnaires are:

- Questionnaire for User Interface Satisfaction (QUIS) (1988)
- Computer System Usability Questionnaire (CSUQ) (1995)
- System Usability Scale (SUS) (1996)

This study answered the questions:

- How well do these questionnaires apply to the assessment of Websites?
- Do any of these questionnaires work well, as an adjunct to a usability test, with relatively small numbers of users?

The results are very interesting. One of the simplest questionnaires studied, SUS (with only 10 rating scales), yielded among the most reliable results across sample sizes. It was also the only one whose questions all address different aspects of the user’s reaction to the website as a whole.

The accuracy of SUS increased very quickly with the number of participants. SUS jumps up to about 75% accuracy at a subject size of 8, and to 100% with a size of 12.

ISO9241-11 Usability Measures covered:

- Effectiveness: Yes
- Efficiency: No
- Satisfaction: Yes

Performance measurement

Performance measurement is an evaluation method used to collect quantitative data about the system in a realistic environment. The method allows comparing with the system and organization requirements and identifying usability issues.

- Quantitative data is collected.
- Qualitative data is not collected.
ISO9241-11 Usability Measures covered:

- Effectiveness: Yes
- Efficiency: Yes
- Satisfaction: No
**Recommended Method: Performance measurement combined with Subjective Assessment**

- Quantitative data is collected.
- Very controlled evaluation method.
- Major usability problems are identified that may not be revealed by less formal testing, including problems related to the specific skills and expectations of the users.
- Evaluation against benchmark data.
- Usability measures: effectiveness, efficiency, satisfaction.
- Emphasis on functionality, and interaction between system and user.

The performance measurement evaluation will be conducted following a well defined testing plan specifying the context of evaluation, the scenarios, tasks and the user groups to be tested. All goals that users are expected to perform need to be quantified to allow measuring the users’ effectiveness, and efficiency. This will allow comparing the results with benchmark data.

The output will be a list of usability problems and summary of usability measures. To complete the effectiveness and efficiency data, a subjective assessment will be performed with each of the users to evaluate the satisfaction level of the user.

**Drawbacks of the method:** A performance measurement does not provide qualitative data.

The proposed solution is to combine the evaluation with a questionnaire as explained above.

**Personnel involved:**
- Usability expert: approx. 1
- Developers
- Users: Approx. 6
Evaluation Plan

This chapter presents the initial plan for evaluating the idSpace tool. The chapter has two main sections – for each of the evaluation phases (described on page 10). For each phase, the evaluation context is first described. Then the chosen scenarios are presented, the actual evaluation procedure outlined and finally, the relevant forms to be used in the evaluation are provided.

**Phase 1**

In the first phase, Prototype V1 which is an early integration of applications is being evaluated. The evaluation aims at providing initial input for the iterative design process that follows. Due to that, no usability experience can be really measured and therefore the evaluation is done by usability experts.

**Context**

Relevant characteristics of the evaluation in this phase are described below:

**Who will evaluate:**
Usability Experts

**Usability Experts:**
As explained in section Phase 1 on page 21, the phase 1 evaluations require usability experts. Currently, only one usability expert is available in the consortium. Therefore, this usability expert, from Space Applications Services, will prepare a usability evaluation course. The course will be given by the usability expert to 4-8 Space Applications Services personnel and out of this group at least 5 trained personnel will be available for conducting the evaluations in the project. For the course syllabus see section 0.

**Evaluation dates:**
- The prototype V1 software should be available by 1-Feb-2009. The evaluation of the first phase is planned for mid February 2009.
- The evaluation report will be released by 30-Feb-2009

**Evaluation platform:**
Since the evaluated software is a web-based software, different platforms will be used to evaluate it. In Space Applications Services, the common practice is that different people use different hardware (Intel based hardware or Apple hardware), and different operating systems (Windows XP, Linux, Mac OS X).

The evaluation by the group of usability experts will be done remotely. That is, each usability expert will evaluate the prototype from a different place, using the network in order to connect to the prototype server.

**The Context of Use:**
In order to help the usability experts focus on the circumstances of actual or intended product usage, a Context of Usage (CoU) analysis should be carried out before the actual evaluation. Such CoU is used as a planning aid for selecting which aspects of the system need to be evaluated, and what are the most meaningful circumstances under which evaluation can take place. This will ensure that the evaluation realistically captures important elements of how the product will be actually used by its end-users.
Three important elements of the CoU are sets of questions which deal with the following issues:

- Who will use the system (Users)
- What will they do with the system (Tasks)
- Where will they use the system (Environment)

The CoU should be available in advance of the actual evaluation to the usability experts.

_Scenarios_

The Phase 1 evaluation will be perform based on a categorization and identification of usability problems based on a set of heuristics. No scenarios shall be developed for this evaluation method.

_Procedure_

The evaluation is performed by having each individual evaluator inspect the system alone. During the evaluation session, the evaluator goes through the system several times and inspects the various dialogue / interface elements and compares them with a list of recognized usability principles (the heuristics). We will use the 10 heuristics by Nielson (Nielsen, 2004) (see Appendix 0) and add some system specific heuristics.

In addition to the checklist of general and category-specific heuristics, the evaluator can further consider any additional usability principles or results that come to mind that may be relevant for any specific system element.

Only after all evaluations have been completed the evaluators are allowed to communicate and have their findings aggregated in a debriefing session. This ensures independent and unbiased evaluations from each evaluator. The results of the evaluation can be recorded either as written reports from each evaluator or by having the evaluators verbalize their comments to an observer as they go through the system.

_Processing the results_

After the evaluation period, the analysts can collate the problem lists and the individual items can be prioritised. The output from using the heuristic evaluation method is a list of usability problems in the interface with references to those usability principles that were violated by the design in each case in the opinion of the evaluator.

A report detailing the identified problems is written and fed back to the development team. The report should clearly define the ranking scheme used if the problem lists have been prioritised.

_The idSpace e-Delphi study_

The proposed e-Delphi study will address a group of circa 15-20 representative experts both from relevant academic and business backgrounds. Invited experts will discuss in three consecutive rounds questions to elucidate their opinions. To make
use of the experts knowledge and their practical experiences we envisage to elicit their ideas on creativity enhancement for idSpace context more in general as well as on requirements for idSpace and finally on usefulness and usability of the prototypical proposition at the time of the Delphi study and suggestions for improvements.

We will use the concepts and design specifications generated in the first and second deliverables plus the first prototype v1 to frame our questions and ask the experts for feedback and improvement suggestions.

This includes that both open and closed questions will be asked about
- learning strategies of experts and novices
- innovation and creativity strategies of experts and novices
- collaboration and sharing
- creativity in workplace settings
- organisational issues related to innovation

The experts will be recruited from:
- known colleague experts working at the idSpace partner institutes
- known academic and business experts from the networks of idSpace members
Especially experts spotted at conferences and idSpace dissemination meetings.

The e-Delphi study is planned early spring. At that moment in time we expect idSpace to have enough substance and prototype v1 contours to benefit not only from the experts views on creativity support for new product design in general (session 1) but also ask for their feedback on the actual state of idSpace requirements specification (session 2) and suggest possible improvements (session 3). Planning the e-Delphi early spring on the other hand ensure that results from the study can still be used for the design of the 2nd prototype.

The OUNL who has expertise in e-Delphi studies will take the lead in the organization and documentation of the e-Delphi study. It will be decided in cooperation with WP4 if the idSpace platform in the actual state can be used as platform for the e-Delphi study.

The Delphi study will consist of three rounds; the second and third round confronting experts with the results of the previous round:

- 1st e-Delphi round/session: expert debate set of by statements to trigger experts to articulate their own vision and ideas about idSpace needs: problems that require idSpace type of solutions, their ideas on creativity support for work/learning, etc.
- 2nd e-Delphi round: using questionnaires focussed on comments /evaluation of the existing idSpace proposition
- 3rd e-Delphi round focussed on opportunities for improvements/alternatives using a mixture: questionnaire based plus open (statement) approach.

Planning:
- 15- January – 20 February Preparation
- Mid February Invitation experts
• +/- 5-20 March e-Delphi execution

Appendix info: Short summary of Delphi method:

• The Delphi method is a systematic interactive knowledge elicitation and forecasting method for obtaining opinions from a panel of independent experts. The carefully selected experts react to statements and questionnaires in two or more rounds. After each round, a facilitator provides an anonymous summary of the experts’ ideas from the previous round as well as the reasons they provided for their judgments. Participants are able to take the input of others into account and to adapt or revise their earlier ideas in light of the replies of other members of the group. The process is stopped after a pre-defined period and stop criterion

• The actual flow for the idSpace e-Delphi study will be described in the scenario.

Phase 2
In the second phase Prototype V2 will be evaluated. The second prototype is the result of the iterative design process and design documentation as well as early User Manual. In this second evaluation, the target is to evaluate the usefulness of the software. The evaluation is done by different user groups and each evaluation session is conducted by a Usability Expert.

Context
The context of the evaluation in this phase is described here (who will evaluate, when, on what platform).

Who will evaluate:

Development Team: a development team which includes 3-5 members from Space Applications Services will be available for evaluating the second prototype. The development team which is likely to be involved is doing work related to knowledge management and natural language processing. It should be noted that this team is not involved in the development of the idSpace tool. In the context of the team’s work, two types of activities might be considered: solving blocking problems, and harvesting ideas for relevant Invitations to Tender (ITTs).

Strategy Team: a strategy team which includes 3-5 members from Space Applications Services will be available for evaluating the second prototype. The strategy team is involved in commercial, strategic and organizational decision-making for examining different commercial directions, and conducts therefore more often than other teams brainstorming sessions.

Students Groups: three groups of 4-6 students each will evaluate the tool. Unlike the development and strategy teams above, the students groups (at least in this stage of their careers) do not really belong to the target group of the tool. Therefore, these groups will especially contribute in determining the effectiveness and efficiency of the tool, while their input related to satisfaction will have a minor effect on the overall assessment.
Evaluation dates:

- The evaluation report will be released by 30-Nov-2009.
- The prototype V2 software should be available by 1-Oct-2009.

Evaluation platform:

The evaluated software is a web-based software, and different platforms may be used to evaluate it. Because the prototype has been already evaluated by different platforms in the first phase, no constraints will be setup on the platforms to be used in the second phase.

The Context of Use:

The Context of Usage (CoU) Analysis derived at the earlier stage of the project will evolve following new available project design and/or evaluation information from the 1st phase.

The updated CoU will drive the formulation of a realistic test plan and associated evaluation scenarios and should therefore be available to the managers of the evaluation campaign well in advance of the preparatory phase of the evaluation.

Scenarios

Different types of evaluation scenarios will be used:

Technical Problems: During the work of development teams, different technical problems may arise. These problems should be problems which cannot be solved by exercising usual techniques, and creative ideas are needed in order to cope with them. A good indicator for such a problem is the fact that usually the solution for the problem is pushing the state of the art in the problem domain.

For example, the problem of expressing events (such as a train arrival) using Topic Maps was dealt with in the past by a development team of Space Applications Services.

Toward the evaluation date, such problems will be collected, and will be used as evaluation scenarios.

Reactions to ITTs: ITTs (Invitation to Tender) often provide good evaluation scenarios. The ITTs define targets and constraints and usually seek for proposals of creative solutions.

For example, FP7 calls or ESA ITTs can be considered valid evaluation scenarios.

Toward the evaluation date, such ITTs will be collected, and will be used during the evaluation.

Strategy questions: Several times a year, strategy directions are examined by the senior personnel in Space Applications Services. Many times, these discussions
include brainstorming to understand how certain technologies can be commercialized.

For example, in the past, the directions taken by Space Applications Services with relation to Topic Maps technology were examined.

Because the strategic discussions include only few and vague constraints, the work is very close to blue sky thinking, and therefore provides good evaluation scenarios.

**Creativity Exercises:** The creativity exercises suppose to challenge the evaluating users with assignments that demand creative thinking.

Examples of such exercises are:

- Invent a new animal.
- Invent a new product to be used at home.
- Suggest a unique event to be carried out in the university.
- Suggest other creative exercises.

The work on the creative exercises is usually done without any realistic relationship to the context in which the tool will be used, but on the other hands, they will provide a good indicator for measuring the creativity output of the evaluation session.

**Procedure**

The major objective of the performance measurement is to provide data on the effectiveness and efficiency of the users' interaction with the system, thus enabling comparisons with similar products, or with previous versions of the product under development. It can also highlight areas where a product can be enhanced to improve usability. The major steps of this method are listed below.

- **Specification of the Context of Evaluation:**
  From the CoU the Context of Evaluation is derived so that the evaluation can be carried out in conditions as close as possible to those in which the system will be used.
Preparing the evaluation:
The evaluation measures the performance of the users as they perform a set of tasks within the context of evaluation and the following defined scenarios. Examples of evaluation measures include:
  o Number and quality of ideas.
  o Assessment of ideas.
  o Appropriate application of creativity techniques.
  o Time to generate successful solutions.
  o Number of features never used by the user.
  o Number of system features that users can remember during the debriefing.
  o Proportion of users stating that they would prefer using the system over some specified benchmark.
Together with the definition of the evaluation measures, the associated measurement instruments (e.g., observation, video recording or questionnaires) will be identified as well. The scenarios as allocated to the dedicated user groups under the section scenarios on page 32 will be further elaborated as part of the preparatory activities. All logistical arrangements should be made with respect to test facilities, user invitation, time frame etc.

A necessary input to the preparation phase is a list of user requirements and design objectives that need to be addressed during the evaluation. The output of the preparatory phase is a Usability Test Plan which details the Evaluation Test Procedures covering all identified requirements.

The Evaluation Test Plan will enclose a compliancy matrix (traceability matrix) mapping the list of user requirements and the project objectives to the test procedures. In addition, the compliancy matrix will detail per test procedure, the expected result.

Such compliance matrix will serve to record the results from the tests and derive deviations, exceptions and limitations compared to the specified objectives. Further, it will testify that each requirement has been fulfilled as per the stated compliance status.

As a baseline for the preparation activities, two days per group will be allocated for the evaluations. Different sessions will be conducted with and without the system to be tested in order to allow comparison with the benchmark data.

**Performing the user tests:**
The different evaluation sessions will be performed according to the defined test procedures detailed during the preparatory step. Measurements are recorded and stored for processing.

In addition to the performance test, Subjective Assessment questionnaires will be distributed among the participants to gather qualitative data to evaluate their satisfaction level (see 0 for the questionnaire).

**Cognitive load theory**
The use of the cognitive load theory is especially relevant in the context of the IdSpace project because it underlines the differences between expert and novices in learning and in using tools such as idSpace. To measure cognitive load, objective measures can be used, such as the users physical state (e.g., using Event Stream software). Also subjective measurements are used frequently. In the context of the idSpace project a subjective measurement is most appropriate given the phase of development of the idSpace prototypes.

To measure cognitive load, a validated questionnaire will be used. The team proposes to use the Nasa Task Load Index, using the lickert scale (see 0).

---

4 The benchmark will be an alternative system/tool that can be choosen by the participants.
Processing the results

The evaluation measurements derived from the test session are used to calculate metrics which provide quantitative measures of usability.

An Evaluation Report will be produced describing the conditions of evaluation, the results from both the Performance Measurement and the Subjective Assessment, and the final recommendations on the project from usability perspective.
Conclusion

This document is the basis for the evaluation activities in idSpace. It suites the needs of the different phases of idSpace and establishes the reasons for the chosen evaluation methods in the different phases.

The planning for the evaluation is formulated as detailed as the current stage of the project allows. The current outline sets the framework, activities and procedures for the evaluation test. The outcome of the preparatory phase (paragraph 0) is a detailed test plan that will serve as the test guide and will be integrated in the overall evaluation plan and the respective evaluation reports.

The current preliminary planning incorporates the system information that is available at the preparation date of this document (e.g., system requirements and design info, outcome and results from the first Evaluation Phase, measurement tools available at facilities, etc.).

For the evaluation of prototype v1 we will use "heuristic evaluation", which is a method for finding both major and minor problems in a user interface design.

With selected scenarios we will analyze the effectiveness and efficiency of the prototype v2 of the idSpace platform. As subcategories of the evaluation measures we will analyze the following features: opportunities for meaningful, goal-oriented communication within the team, problem definition, group composition, goal setting (e.g., prime objective), provision of information (e.g., learning from each other, about creativity techniques), provision of appropriate creativity techniques, number and quality of ideas, understanding of ideas from the group, assessment of ideas (weighting of solutions), compiling minutes, ranking of ideas.
References


Appendix

Questionnaire SUS - System Usability Scale

Using SUS

The SU scale is generally used after the participant has had an opportunity to use the system being evaluated, but before any debriefing or discussion takes place. Participant should be asked to record their immediate response to each item, rather than thinking about items for a long time. All items should be checked. If a participant feels that they cannot respond to a particular item, they should mark the centre point of the scale.

Scoring SUS

SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own.

To calculate the SUS score, first sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1, 3, 5, 7 and 9 the score contribution is the scale position minus 1. For items 2, 4, 6, 8 and 10, the contribution is 5 minus the scale position.

Multiply the sum of the scores by 2.5 to obtain the overall value of SU. SUS scores have a range of 0 to 100.
D5.2 Evaluation plan & planning

1. I think that I would like to use this system frequently.
   *Strongly disagree* [ ] [ ] [ ] [ ] [ ] *Strongly agree* [ ] [ ] [ ] [ ] [ ]
   1 2 3 4 5

2. I found the system unnecessarily complex.
   *Strongly disagree* [ ] [ ] [ ] [ ] [ ] *Strongly agree* [ ] [ ] [ ] [ ] [ ]
   1 2 3 4 5

3. I thought the system was easy to use.
   *Strongly disagree* [ ] [ ] [ ] [ ] [ ] *Strongly agree* [ ] [ ] [ ] [ ] [ ]
   1 2 3 4 5

4. I think that I would need the support of a technical person to be able to use this system.
   *Strongly disagree* [ ] [ ] [ ] [ ] [ ] *Strongly agree* [ ] [ ] [ ] [ ] [ ]
   1 2 3 4 5

5. I found the various functions in this system were well integrated.
   *Strongly disagree* [ ] [ ] [ ] [ ] [ ] *Strongly agree* [ ] [ ] [ ] [ ] [ ]
   1 2 3 4 5

6. I thought there was too much inconsistency in this system.
   *Strongly disagree* [ ] [ ] [ ] [ ] [ ] *Strongly agree* [ ] [ ] [ ] [ ] [ ]
   1 2 3 4 5

7. I would imagine that most people would learn to use this system very quickly.
   *Strongly disagree* [ ] [ ] [ ] [ ] [ ] *Strongly agree* [ ] [ ] [ ] [ ] [ ]
   1 2 3 4 5

8. I found the system very cumbersome to use.
   *Strongly disagree* [ ] [ ] [ ] [ ] [ ] *Strongly agree* [ ] [ ] [ ] [ ] [ ]
   1 2 3 4 5

9. I felt very confident using the system.
   *Strongly disagree* [ ] [ ] [ ] [ ] [ ] *Strongly agree* [ ] [ ] [ ] [ ] [ ]
   1 2 3 4 5

10. I needed to learn a lot of things before I could get going with this system.
    *Strongly disagree* [ ] [ ] [ ] [ ] [ ] *Strongly agree* [ ] [ ] [ ] [ ] [ ]
    1 2 3 4 5

Fig. 1: System Usability Scale (1996.)
**Usability Heuristics by Nielsen (Nielsen, 1994)**

**Visibility of system status**
The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

**Match between system and the real world**
The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

**User control and freedom**
Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

**Consistency and standards**
Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

**Error prevention**
Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

**Recognition rather than recall**
Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

**Flexibility and efficiency of use**
Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

**Aesthetic and minimalist design**
Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

**Help users recognize, diagnose, and recover from errors**
Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

**Help and documentation**
Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.
### NASA TLX Paper and Pencil Version

**NASA Task Load Index**

Hart and Staveland’s NASA Task Load Index (TLX) method assesses workload on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

<table>
<thead>
<tr>
<th>Name</th>
<th>Task</th>
<th>Date</th>
</tr>
</thead>
</table>

**Mental Demand** How mentally demanding was the task?

| Very Low | Very High |

**Physical Demand** How physically demanding was the task?

| Very Low | Very High |

**Temporal Demand** How hurried or rushed was the pace of the task?

| Very Low | Very High |

**Performance** How successful were you in accomplishing what you were asked to do?

| Perfect | Failure |

**Effort** How hard did you have to work to accomplish your level of performance?

| Very Low | Very High |

**Frustration** How insecure, discouraged, irritated, stressed, and annoyed were you?

| Very Low | Very High |
## Usability Training - Course Syllabus

### Training: INTRODUCTION TO USABILITY EVALUATION

<table>
<thead>
<tr>
<th>1. DATES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,00 Days for Training Course – 16th February 2009</td>
</tr>
<tr>
<td>1,00 Days for Usability Evaluation exercise – 17th February 2009</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. OBJECTIVES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject course aims to provide an overview of the common tools and techniques available to assess usability of an interface or system, and to provide guidelines on how to successfully perform these evaluations.</td>
</tr>
<tr>
<td>The course will explore the difference between expert reviews and usability testing, which of the evaluations is right based upon given situations, how to plan and conduct those evaluations, and finally how to analyze and report back on the findings.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. COURSE CAPACITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. 8 participants</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. TARGET POPULATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>System designers</td>
</tr>
<tr>
<td>Project managers</td>
</tr>
<tr>
<td>Testers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. TRAINEES PRE-REQUISITES</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6. LANGUAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7. LOCATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Application Services, Zaventem</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>8. DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>The course will provide the participants with a range of common evaluation techniques for usability issues at every stage of the development.</td>
</tr>
<tr>
<td>Special attention is given to heuristic evaluation and performance testing to suit the needs of the idSpace project.</td>
</tr>
<tr>
<td>The second day of the course will provide hands-on instruction applying the theory learned during the first day. This will include an introduction into the idSpace project, general overview of the purpose and functionality of the idSpace system, evaluation of the system and debriefing.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. DOCUMENTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powerpoint slides</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10. GLOBAL FOOTPRINT</th>
</tr>
</thead>
</table>
### Day 01

<table>
<thead>
<tr>
<th>Subject</th>
<th>Time</th>
<th>Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>2hrs</td>
<td>Classroom instruction</td>
</tr>
<tr>
<td>• About usability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Overview of key usability principles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Usability evaluation methods and techniques</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert Reviews – Usability Inspection Methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Heuristic Evaluation Process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Example</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Alternative methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usability Testing</td>
<td>2hrs</td>
<td>Classroom instruction</td>
</tr>
<tr>
<td>• Types of testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Planning and co-ordinating the testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Analysis and Reporting</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Day 02

<table>
<thead>
<tr>
<th>Subject</th>
<th>Time</th>
<th>Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to idSpace Project</td>
<td>30min</td>
<td>Classroom instruction</td>
</tr>
<tr>
<td>• Scope</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Overview</td>
<td>1hrs</td>
<td>IdSpace system</td>
</tr>
<tr>
<td>• Objectives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Functionality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td>2hrs</td>
<td>IdSpace system</td>
</tr>
<tr>
<td>Debriefing</td>
<td>1hrs</td>
<td>Group discussion</td>
</tr>
<tr>
<td>Conclusion and wrap-up</td>
<td>1hrs</td>
<td>Group discussion</td>
</tr>
</tbody>
</table>