MOSAICA

Semantically Enhanced, Multfaceted, Collaborative Access to Cultural Heritage

Project IST-034984

D6.1

Validation and evaluation plan

Editors: TECH, SHU
Contributors: TECH, SHU, ORT
Dissemination Level: **Public**
Date: 2-July-2008

Executive Summary

The aim of this deliverable is to create a plan to carry out an effective assessment of the project outcomes. Part 1 of this document has been divided into two main sections. The first one contains details of the validation process. It details: the validation framework; use cases and user groups; the system functionality in terms of system features; and the architecture in terms of the layers deployed. The second section contains evaluation plan and is divided into three main sections: Research Methodology, Societal Impact and Multi-Cultural Evaluation. The content in Part 2 is comprised of basics of evaluation tools and their detail description.

CONTENTS

1. Introduction 7

2. Validation Framework 8

2.1 requirements Management Process 10

2.2. MOSICA System Validation 13

2.2.1 Use Case and User Groups 13

2.2.2 Frontend Components 16

2.2.2.1 Semantic Search and Browsing 17

2.2.2.2 Semantic Annotator 18

2.2.2.3 GIS/TIS 19

2.2.2.4 Virtual Expedition Player 20

2.2.2.5 Virtual Expedition Editor 21

2.2.2.6 Online Ontology Editor 22

2.2.2.7 MOSICA Semantic Layer 22

2.2.3 Logistical Plan 23

2.2.3.1 Validation Plan Overview 23

2.2.3.2 Main System Functionalities 25

2.2.3.3 Timeline of Validation 25

2.2.3.4 Users Validation Groups.....25

3. Evaluation Framework 28

3.1. Evaluation Research Methods 30

3.2. Quantitative and Qualitative Research and Tools 32

3.2.1 Online Questionnaires and Surveys 34

3.2.2 Online Knowledge Sharing Systems.....36

3.2.3 Semi-structured Interviews 37

3.2.4 Access Log Analysis 38

3.3 Social Impact Evaluation 39

3.3.1 Open- Mindedness 39

3.3.1.1 Development and validation of Open-Mindedness Questionnaire 41

3.3.2 Usability 41

3.3.2.1 Development and Validation of Usability and Learning Outcomes Questionnaire 44

3.3.2.2 Development and Validation of Semi-Structured Interview Questions 44
3.3.3 Expedition Diary 45
3.3.3.1 Expedition Diary Structure 45
3.3.3.2 Expedition Diary Design 45
3.3.4 A summary of the Social Impact Evaluation Tools 47
3.3.5 A logistical Plan for Societal Impact Evaluation 50
3.4 Multi-Cultural Evaluation 51
3.4.1. Understanding Cross-cultural and Multi-cultural aspects of Evaluation [13] 51
3.4.1.1 Definitions 51
3.4.1.2 Understanding the Influence of Culture and Evaluation Practice 52
3.4.1.3 Defining Cultural Competence within an Evaluation Context 52
3.4.1.4 Cultural sensitivity and Evaluation 54
3.4.1.5 Defining Cultural Competence: Evaluation in Multi-cultural settings 56
3.4.1.6 Exploring the role of assumptions in Evaluating Practice 57
3.4.2 Guiding principles 58
3.4.2.1 Inclusion in Design and Implementation 58
3.4.2.2 Appropriate measures and success 60
3.4.2.3 Cultural and system Analysis 61
3.4.3 A holistic multi-cultural Evaluation Framework for MOSAICA 62
3.4.4 Logistical Plan for Multi-cultural Evaluation 64
4 References 67
5. Appendices 72
Appendix A - The open-mindedness questionnaire 73
Appendix B - Usability and Educational outcomes Questionnaire 75
Appendix C - Semi-structured Interview Questions 77
Appendix D - Open-ended questionnaire - 1 for Holistic Evaluation 79
Appendix E - Open-ended questionnaire - 2 for Holistic Evaluation 80

LIST OF FIGURES

Figure 1. MOSAICA Validation Activities [8] 9

Figure 2. MOSAICA Validation Framework 9

Figure 3. WP8 Project Management 11

Figure 4. Use case and user groups 15

Figure 5. Validation plan overview 23

Figure 6. First validation 24

Figure 7. Second validation 24

Figure 8. Deductive Vs Inductive Reasoning 30

Figure 9. MOSAICA social impact, usability and learning outcomes evaluation.....50

LIST OF TABLES

Table 1. Validation criteria, measurement methods and expected results of semantic search and browsing 18

Table 2. Validation criteria, measurement methods and expected results of semantic annotator 19

Table 3. Validation criteria, measurement methods and expected results of GIS/TIS 20

Table 4. Validation criteria, measurement methods and expected results of virtual expedition player 21

Table 5. Validation criteria, measurement methods and expected results of virtual expedition editor 22

Table 6. Validation criteria, measurement methods and expected results of Online Ontology Editor 22

Table 7. Validation criteria, measurement methods and expected results of Semantic Layer 23

Table 8. The profile of user groups within the project 25

Table 9. The profile of user groups outside the project.....27

Table 10. Evaluation tools to be integrated as part of MOSAICA platform.....48

Table 11. MCE: Evaluators of targeted Focus Groups 64

Table 12. Participants for Evaluating Cultural Heritage Ontology 65

Table 13. Pilot Study Participants 65

Table 14. Expert Study Participants 66

1. INTRODUCTION

The objective of creating a validation and evaluation plan is to carry out an effective assessment of the project outcomes. This is necessary for evaluating and validating the project activity and corresponding results from several distinct perspectives such as individual technological components (verification), as well as the entire system (validation) being developed in the project, track the progress in development and research based on system requirements, systematic scientific insight into expected impact of MOSAICA's usage on the modification of attitudes and behaviour among the users and finally an impartial evaluation of the use of MOSAICA's approach and technologies in additional, diversified cultural contexts. To that aim, in this document we will draw our plan to establish and conduct validation and evaluation from three different perspectives: System Validation, Societal Impact Evaluation and Multi-Cultural Evaluation. The proposed plan for Validation & Evaluation has been developed based on a validation framework that comprises of activities for System Validation and an evaluation framework for Societal Impact and Multi-Cultural Evaluation. In the subsequent sections we will elaborate on each of these frameworks. First we provide some background material on validation frameworks.

2. VALIDATION FRAMEWORK

A significant body of knowledge (ISO 8402:1994 [1], European Commission [2], [3], IEEE [4],[5], ISO [6], et al) considers “verification” and “validation” as separate and distinct terms. Some use the terms "verification" and "validation" interchangeably, or in some cases refer to software "verification, validation, and testing (VV&T)" as if it is a single paradigm, with no distinction among the three terms. Here verification includes activities involved in verifying the feasibility and functionalities of a software deliverable through utilization of a small but sufficient sample of users in a real-life situation to yield preliminary findings on user acceptance (See European Commission 1994b, [2]).

WP7 builds on this body of knowledge (EU Commission [2], IEEE [4],[5], ISO [6], et. al) that considers verification and validation to be distinctly different activities within the development lifecycle. Verification is the provision of objective evidence that the design outputs of a particular phase of the MOSAICA system development life cycle meets all the specified requirements for that phase. Validation is considered as the confirmation by examination and provision of objective evidence that software specifications conform to user needs and its intended uses, and that the particular user requirements [7] implemented through software can be consistently fulfilled. Accordingly, verification embodies consistency, completeness, and correctness of a given MOSAICA individual component and its supporting documentation, as it is developed, and provides the input for the subsequent validation activity for same.

Consequently, all validation activities are highly dependent upon comprehensive testing, inspections, analyses, and other verification tasks performed at each stage of the MOSAIC system development life cycle as graphically illustrated in figure 1.

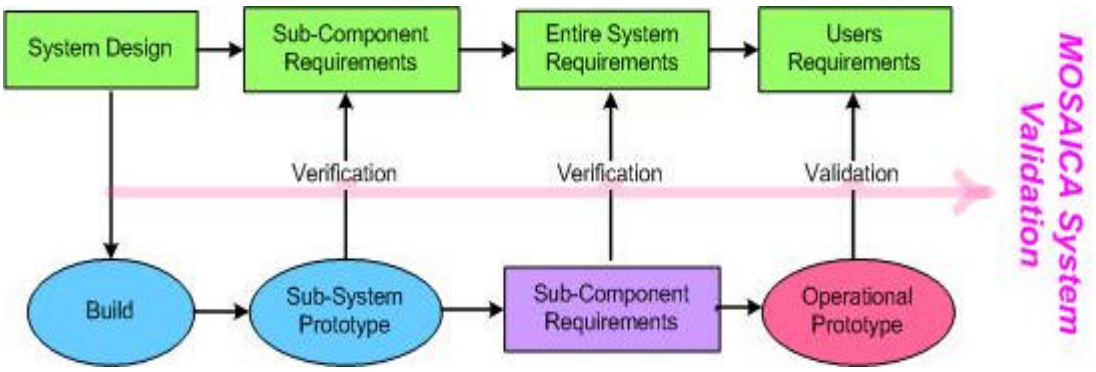


Figure 1. MOSAICA Validation Activities [8]

To this end the project validation activities have been conducted during the MOSAICA system development phases to ensure that all requirements have been fulfilled, as well as at the end of the MOSAICA development cycle which will constitute as the final evaluation.

Clearly, verification and validation activities can not be absolute since unlimited testing by the developers is simply not feasible and on the same token it is difficult to discern just how much evidence is constituted as sufficient to validate an individual component. Accordingly, the pragmatic strategy for the validation activities in large measure is a matter of developing a “level of confidence” that the MOSAICA system meets all requirements and user expectations set out in the MOSAICA Requirements Documents - Deliverable D5.1.

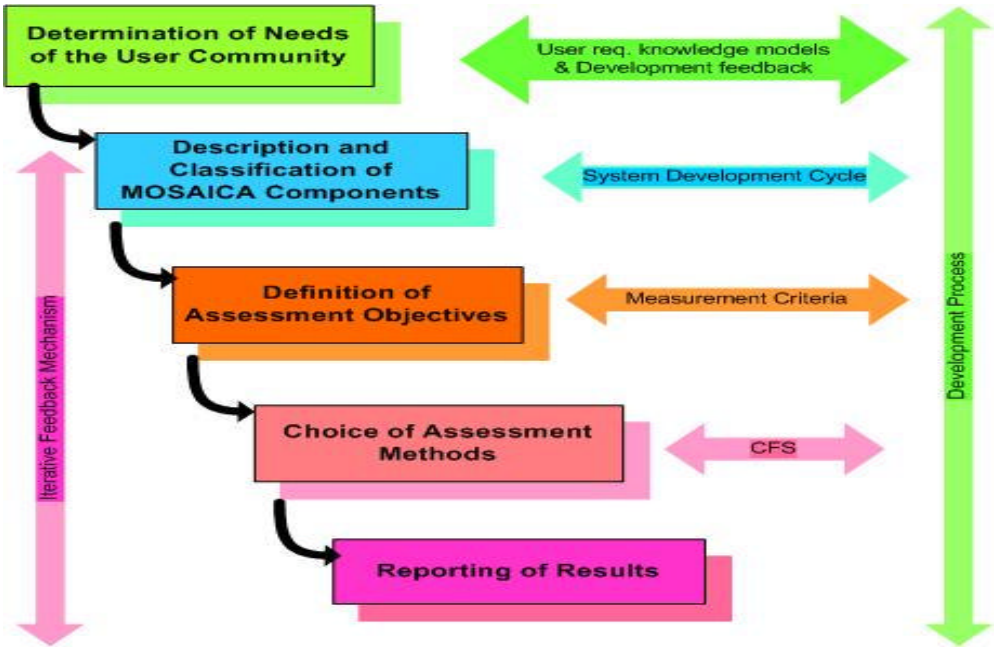


Figure 2. MOSAICA Validation Framework

Figure 2 graphically illustrates the MOSAICA validation framework that WP1, WP2, WP3, WP4, WP5 and WP6 will be engaged upon, demonstrating the logical relationship between the different stages of validation activities of MOSAICA system. It should be noted that the evaluation and validation procedures and their execution is based upon Goal Directed Project Management Methodology [9,10].

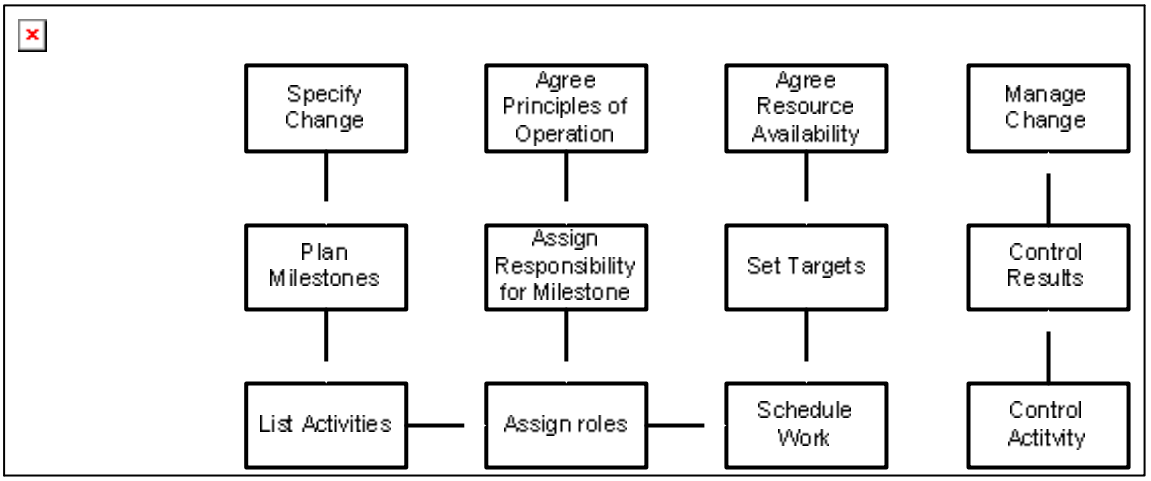
2.1 Requirements Management Process

The overall management processes for the conduct of the WP1, WP2, WP3, WP4 and WP5 work packages include the planning, organization, coordination, control and management in relation to ***validation and evaluation process*** (WP6). Principle methodology utilized to manage these is based on Goal Directed Project Management methodology (GDPM) [10]. The methodology encourages a team oriented approach towards planning and controlling projects.

Building on the above, the overall work package plan consists of several hierarchical levels addressing:

1. The global; that is, the management of the overall objectives of the work package and the management of the relationship between WP1 and other work packages.
2. The administrative; that is, the management of the administrative tasks within the work package including the deliverables of same.
3. The operational; that is, the management of the daily operations during the life of the work package.

Figure 3 illustrates the interrelationship between planning, organisation, co-ordination, and control and the global, administrative, and operational layers noted.


Specify Change
Agree Principles of Operation
Agree Resource Availability
Manage Change
Plan Milestones
Assign Responsibility for Milestone
Set Targets
Control Results
List Activities
Assign roles
Schedule Work
Control Actitvity

To achieve this control and transparency the following high level processes have been created:

1. The user and development communities review the allocated project requirements in D5.1 before they evaluate the system.
2. The development community utilise the allocated requirements as the basis for development plans, work products, and activities in context of MOSAICA development life cycle.
3. Changes to the allocated requirements are reviewed and incorporated into the system.

Specifically in order to execute the above processes the following procedures have been put in place, to discuss issues noted above.

1. Tele-conference meetings involving the validation and verification group (every week)
2. E-mail communication
3. Joint briefing sessions with external user groups

In conduct of the procedures noted above, the following matrix is used to document and control the process:

Requirements Control	Control Actions
What is the status of each requirement?	All the requirements documented, individually controlled and assessed.
What is the rate of change for the allocated requirements?	Maintaining and logging all changes to the requirements. These changes can then be compared to a preliminary benchmark.
What is the cause of the change to the allocated requirements?	Re-visit requirement gathering process / consult user community
How will requirement changes affect the development process?	The impact of each and every change to the requirements (Requirements Integrity [RI]).
Does the MOSAICA system satisfy the requirements?	Continually determining the functionality of the system specifications and resulting software and comparing it against the RI.

The above matrix serves as an internal tracking mechanism for requirements tracking and management process of the MOSAICA project.

In order to maintain the high level of accuracy for the requirements tracking process, Savannah [11] issue tracking system has been implemented. This takes advantage of dynamic issue management strategy; committing, tracking and modifying issues. Utilizing issue management features such as:

1. **Relationships:** Tracking of relationships between issues. Dependencies between issues are flagged and tracked through either the issue entry process or while reviewing one or multiple issues.
2. **Change tracking:** All changes made along with information about the user who instigated the change, when a change is made and any comments committed along with the change are tracked with the issue.
3. **Comment viewing and editing:** All comments made to the issue are tracked. Comments allow support or other type conversations to be viewable and editable by any parties interested in the progress of an issue.

2.2 MOSAICA System Validation

2.2.1. *Use Case & User Groups*

MOSAICA requirements outline three distinct use scenarios:

1. *Explorative usage.* – Users will be able to visit places that evoke their interest and motivation, by merely zooming-in on a particular area on MOSAICA's geographical maps of Europe, or by exploring MOSAICA semantic directory, or by submitting a query. They can conduct, and record their own pre-planed or spontaneous “expeditions”, and choose places they would like to visit. The users are free to travel the virtual Worlds, and learn about the Jewish heritage in their own time, and as often as they wish.
2. *Guided usage.* – MOSAICA will offer a variety of recommended Virtual Expeditions divided by topics and objectives. Each Virtual Expedition can start

and end at a different location, have a differing time span, and include the display of selected artefacts that are considered to be a "must" within a certain location (e.g., museum, synagogue, or cemetery).

3. *Collaborative usage.* - Educational personnel and the individual users will be able to share their cultural assets and knowledge in two ways: (1) users will be able to annotate exposed cultural objects either with their free-text comments, or semantically annotate them using the MOSAICA Ontology, and to directly contribute content, including photos of artefacts owned by them, that can otherwise not be publicly exposed, and thus exhibit them online, and (2) users can design and record their own Virtual Expeditions, and suggest them to other visitors by storing them in the Repository of Educational Resources.

To complement these scenarios we identify the relevant user groups involved in using the system as stakeholders.

User groups involved in validating the MOSAICA system may be divided into two main types - “User Groups” within the Project and “User Groups” outside the Project involved in validating MOSAICA project recognizing the role and contribution of the latter in independent validation and assessment of the MOSAICA system functionalities.

Members of internal user groups will in charge of the first validation and during the second validation they will act as appraisal managers and liaise with external user groups acting as the interface to the project.

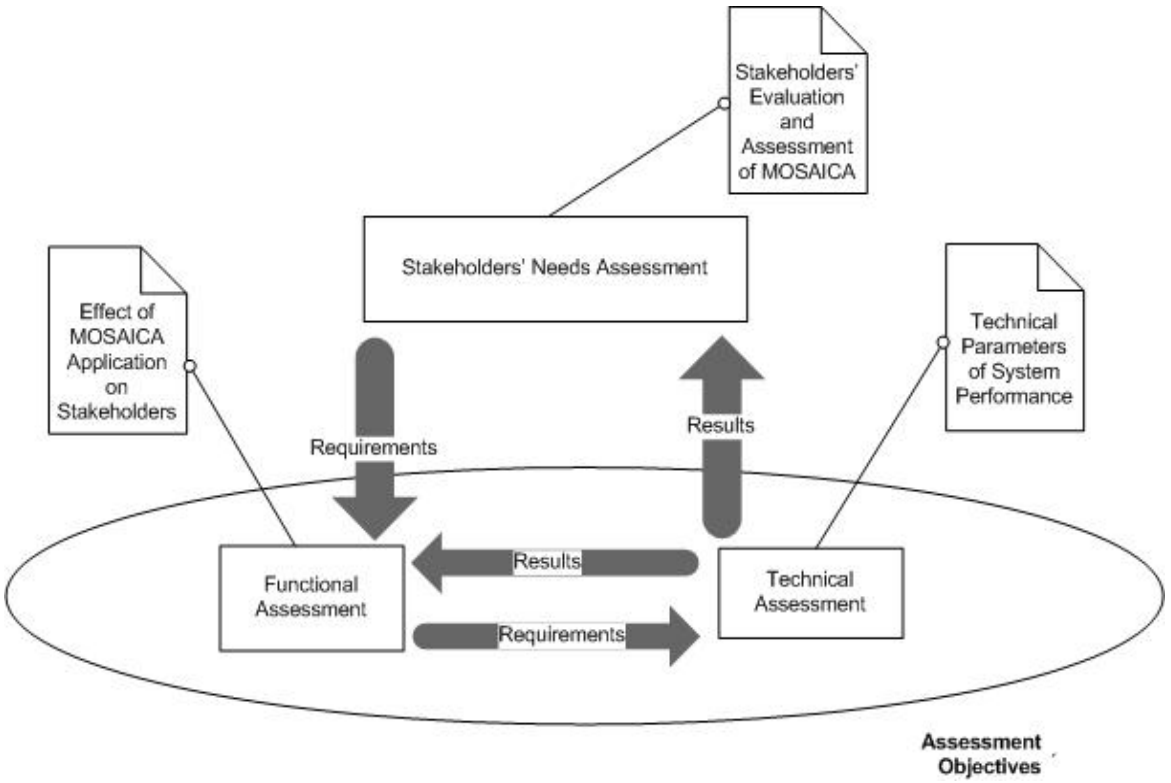


Figure 4. Use case and user groups

Figure 4, illustrates the stakeholder needs in the context of the above closely related to the implementation and use of the developed applications.

All facts and figures given in this section are the result of provisional conceptual planning based on the evaluation frameworks presented. They are subjected to change during the iterative process of evaluation.

The following external user community selections and configuration is a significant activity for the MOSAICA project. MOSAICA system should address the main aspects of promoting multiculturalism through its **explorative**, **collaborative** interfaces and through **guided** facilities. There are interrelations between each of these MOSAICA features and developing tolerant attitudes in a user as well as in the community of users

Based on these scenarios and stakeholders a set of system functionalities and a system architecture was developed.

2.2.2. *Frontend Components*

MOSAICA is an advanced Web-based portal dedicated to the Jewish cultural heritage in Europe, and featuring (1) multifaceted interfaces for knowledge-based exploration and discovery, and (2) online utilities empowering users to collaboratively author and manage cultural resources in the globally distributed environment and must be evaluated by the user community to ensure that the system is successful in all of the required user functionalities.

To ensure subjective evaluation of the MOSAICA system, the presentation and hands on demonstration incorporates clear explanation of project objectives and efforts towards meeting the mandatory deliverables of the project:

GIS (Geographical Information System) empowered map. Rather then hard-coding GIS data, it is our intention to empower the GIS with the Semantic Web technologies, and thus allow semantic references between the information displayed on the map to be created dynamically and online. By applying logical reasoning we intend to enrich the spatial display with the timeline, thus extending the spatial navigation into the temporal dimension.

Semantic directory. Instead consisting of a single, predefined ontology, the MOSAICA Ontology will be generated dynamically and online from multiple, distributed ontologies on Jewish heritage already available over the WWW. Unlike regular directories that are hierarchically organised, MOSAICA semantic directory will allow conceptual navigation by following additional, complex semantic relations.

Semantic search engine. Search and retrieval facility will be empowered with logical reasoning allowing semantic inference, and extending the keyword querying into the domain of conceptualisation. Both semantic directory and semantic search engine will allow integrated access to the diversified (in both content and format) cultural resources by seamlessly aggregating them into a single conceptual frame.

Online semantic annotator. This utility enables users to conceptually model semantic relations between complex and diversified cultural objects and resources. It will provide (1) the facility for free-text annotation allowing users to comment and recommend individual cultural objects, or to instruct other visitors regarding how to

use them, as well as (2) the full integration of MOSAICA ontology enabling users to associate cultural objects with relevant ontological concepts and semantic relations.

Online ontology editor. This utility empowers users with the user-friendly means to further develop their own conceptual models, and dynamically enrich and extend MOSAICA ontology with their own semantics. It will provide two complementary interfaces:

1- for regular users and domain experts a simple interface for editing N-Triple axioms will be provided, while

2- for ontology engineers a full-fledged OWL editor will be integrated.

Virtual Expedition maker. This is actually a set of complementary online utilities that guide users how to design their own Virtual Expeditions, record them, and enhance them with educational instructions and additional activities, such as treasure hunt, portfolio preparation, etc. The design instructions for Virtual Expeditions, and the corresponding utilities will be based on a comprehensive methodology for developing educational resources that will be researched and developed within the project's lifetime.

2.2.2.1. Semantic Search and Browsing

Semantic Search and Browsing (SSB) is the chief graphic user interface to the MOSAICA content. MOSAICA content comprises MOSAICA knowledgebase (which in turn consists of the MOSAICA ontologies and MOSAICA knowledge expressed as assertions) and MOSAICA annotations (which in turn include annotating assertions, references to external content resources, references to internal content resources, as well as the actual internal content resources). SSB provides access to the MOSAICA content through the MOSAICA knowledgebase. Hereby, MOSAICA ontologies can be browsed like WWW directories or digital library catalogues. In addition, MOSAICA ontologies are utilized in the semantic search to formulate high-precision queries executed over the MOSAICA knowledgebase. Finally, regular keyword search enables querying MOSAICA keyword indices including inverted indices of textual content resources, as well as string metadata fields.

SSB is implemented as a Java applet. The frontend runs in a browser, and requires a

minimal installation: the latest version of Java, and potentially additional plug-ins for specific media, if preview of such media is desired. The backend is the Semantic

SSB is the major graphic user interface component and its critical success factor is usability. However, the usability of the SSB depends on other components that cooperate with the SSB, in order to provide the unique user experience.

Table 1. Validation criteria, measurement methods and expected results of semantic search and browsing

Validation criteria	**Measurement method**	**Expected results**
Precision and recall of information retrieval (usability of the high-precision search)	Precision and recall computation on a preset test corpus (a subset of MOSAICA annotations)	Approaching 100% with no false negatives (and depending on the complexity of the expected reasoning[1])
Ranking strategy of information retrieval (usability of both high-precision and keyword search)	User evaluation based on the familiarity with the test corpus. Measurement should be qualitative[2] applying chi-square or similar coefficient of correlation.	High satisfaction (statistically significant high correlation between users' and system's ranking).
Browsing simplicity (usability of browsing facilities)	Time required to find the location of a test concept in the hierarchical display[3]	Browsing time measured in tens of seconds

2.2.2.2. Semantic Annotator

Semantic Annotator (SA) is the system component designed to enable the end user to generate annotations, and attach them to a resource. These annotations are metadata

[1] The test corpus may expect reasoning method that is in fact not implemented, which would lead to reduced precision.
[2] Users could be requested to re-order the corpus in the way that reflects their ranking preferences, and the results could be correlated.
[3] Since the same browsing facilities are also provided in the Semantic Annotator, the usability of the

containing references to the domain ontology. In addition to the generation of annotations, SA enables user to upload locally resident resources onto the P2P network, and thus make them available also to other users.

SA is implemented as a Java applet. The frontend runs in a browser, and requires a minimal installation: the latest version of Java, and potentially additional plug-ins for specific media, if preview of such media is desired. The backend is the Semantic Layer detailed below.

The critical success factor in the development of the SA is the usability. As detailed in the deliverable D2.2 Semantic Annotator, SA was designed following the current usability standards in the so-called *Web 2.0* set for the so-called *social tagging*. The high level of usability is considered to be a pre-condition for recruiting the user's effort to annotate resources for the use of others.

Table 2. Validation criteria, measurement methods and expected results of semantic annotator

Validation criteria	Measurement method	Expected results
Easiness of use (usability)	Average time spent on annotating a resource User feedback	Annotating time measured in tens of seconds High-level of satisfaction with the usability

2.2.2.3. GIS/TIS

GIS is a system of hardware and software used for storage, retrieval, mapping, and analysis of geographic data. Practitioners also regard the total GIS as including the operating personnel and the data that go into the system. Spatial features are stored in a coordinate system (latitude/longitude, Lambert, UTM, etc.), which references a particular place on the earth. Descriptive attributes in tabular form are associated with spatial features. Spatial data and associated attributes in the same coordinate system can then be layered together for mapping and analysis. GIS can be used for scientific investigations, resource management, and development planning.

Semantic Annotator critically depends on this measurement too.

GIS differs from CAD and other graphical computer applications in that all spatial data is geographically referenced to a map projection in an earth coordinate system.

For the most part, spatial data can be "re-projected" from one coordinate system into another, thus data from various sources can be brought together into a common database and integrated using GIS software. Boundaries of spatial features should "register" or align properly when re-projected into the same coordinate system. The fundamental components of spatial data in a GIS are points, lines (arcs), and polygons.

The GIS/TIS has been developed in Java and Flash, a MySQL database is also used to store information about a map.

Table 3. Validation criteria, measurement methods and expected results of GIS/TIS

Validation criteria	Measurement method	Expected results
Easiness of use (simple menu, better view of the resources)	Have a clear view of the resources located on a map	Access needed information quickly

2.2.2.4. Virtual Expedition Player

Virtual Expedition (VE) is a new presentation paradigm for cultural heritage. Virtual Expedition Player (VEP) is the component that enables execution of the VEs. If VE is compared to the HTML, then VEP would be the VE browser.

VEP is just a downscaled version of the VEE Whiteboard component. Therefore, for the details of the implementation see the next chapter.

The critical success factor of the VEP is its ability to provide complex VEs that successfully meet the requirements defined in the D4.1 Virtual Expedition Methodology. Since VEs are based on the storytelling approach, the presentations capabilities of the VEP must provide the implementation of the VE design principles (see 5.1 Design Principles of the D4.1 Virtual Expedition Methodology).

Table 4. Validation criteria, measurement methods and expected results of virtual expedition player

Validation criteria	Measurement method	Expected results
Design principles specified in D4.1 Virtual Expedition Methodology	Face value validation of the implementation[4]	All design principles found their implementation in the technology of choice

2.2.2.5. Virtual Expedition Editor

Virtual Expedition Editor (VEE) is the utility application designed to generate Virtual Expeditions (VE) using the semantically annotated resources. The output of the VEE is used as the input for the VEP.

VEE frontend is implemented as the browser-based HTML and Java Script application developed in Google Web Toolkit (GWT). The backend is Apache Tomcat Server operating as Java servlet server. Frontend and backend interoperate through asynchronous Remote Procedure Calls (RPC). Since VEs require two-way communication (for example, in order to support Real Time Dialog Control) the Comet paradigm, implemented as Rocket (integrated into GWT) was utilized, in order to provide a constant stream of data from server to the client browser. Furthermore, in order to facilitate various kinds of data multiplexer/demultiplexer mechanisms were implemented on both client and server side.

The critical success factor of the VEE is the easiness by which VEs are created and the extent of the creativity that is enabled by the VEE.

[4] Most of the design principles could be implemented in various ways – depending on the underlying technology (HTML/JScript, 3D gaming platform, etc.). The face validation should ensure that the implementation in the technology of choice indeed meets the requirements corresponding to different design principles.

Table 5. Validation criteria, measurement methods and expected results of virtual expedition editor

Validation criteria	Measurement method	Expected results
Easiness of use (usability)	User feedback	High-level of satisfaction with the usability
Support for creativity	User feedback	High-level of satisfaction with the usability

2.2.2.6. Online Ontology Editor

Online (Simple) Ontology Editor (SOE) is a utility application allowing editing of the MOSAICA knowledgebase. SOE is implemented as a Java applet. The frontend runs in a browser, and requires only installation of the latest version of Java The backend is the Semantic Layer detailed below.

The critical success factor for the SOE is the usability factor determining how efficiently a layman can edit MOSAICA knowledgebase.

Table 6. Validation criteria, measurement methods and expected results of Online Ontology Editor

Validation criteria	Measurement method	Expected results
Usability of knowledgebase editing	User feedback	High-level of satisfaction with the usability

2.2.2.7. MOSAICA Semantic Layer

Semantic Layer (SL) is the backend subsystem that provides access to the MOSAICA knowledgebase. SL provides interface to other system components that enables them to obtain on demand data required for execution of their functionalities. SL does not provide merely the access to the row data in the MOSAICA knowledgebase, but applies machine reasoning to the row data, in order to provide logically inferred data.

SL is implemented in Java as a server with Web Services interface.

In contrast to the frontend components that all have various aspects of usability as

their critical success factors, SL's success factor is its ability to be efficiently integrated into the entire system, so that all other components depending on it can function properly.

Table 7. Validation criteria, measurement methods and expected results of Semantic Layer

Validation criteria	Measurement method	Expected results
Scalability	Number of concurrent users	SL is expected to scale up as a Web Server
Quality of Service (QoS)	Percentage of incomplete interactions	SL is expected to meet the highest standards of QoS

2.2.3. *Logistical Plan*

2.2.3.1. Validation Plan Overview

- Cycle of testing sessions
- Usability testing methods
- Finding problems of the prototype
- Continuous improvement

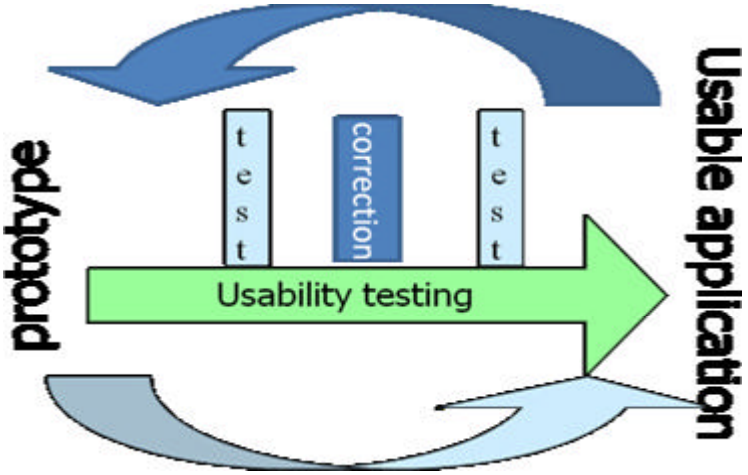


Figure 5. Validation plan overview

First validation

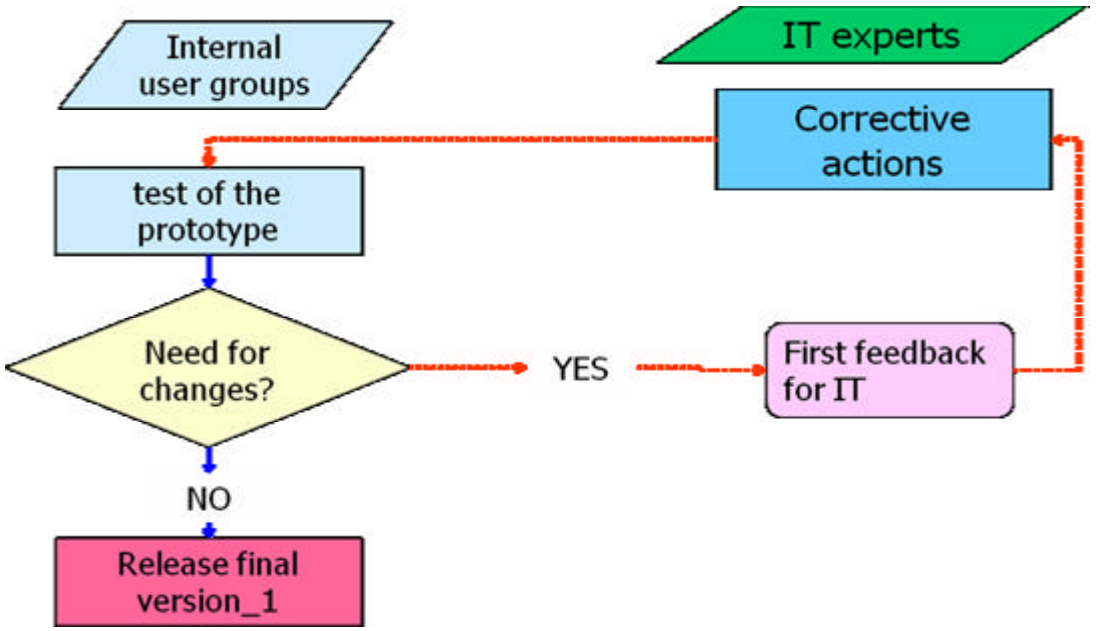


Figure 6. First validation

Second validation

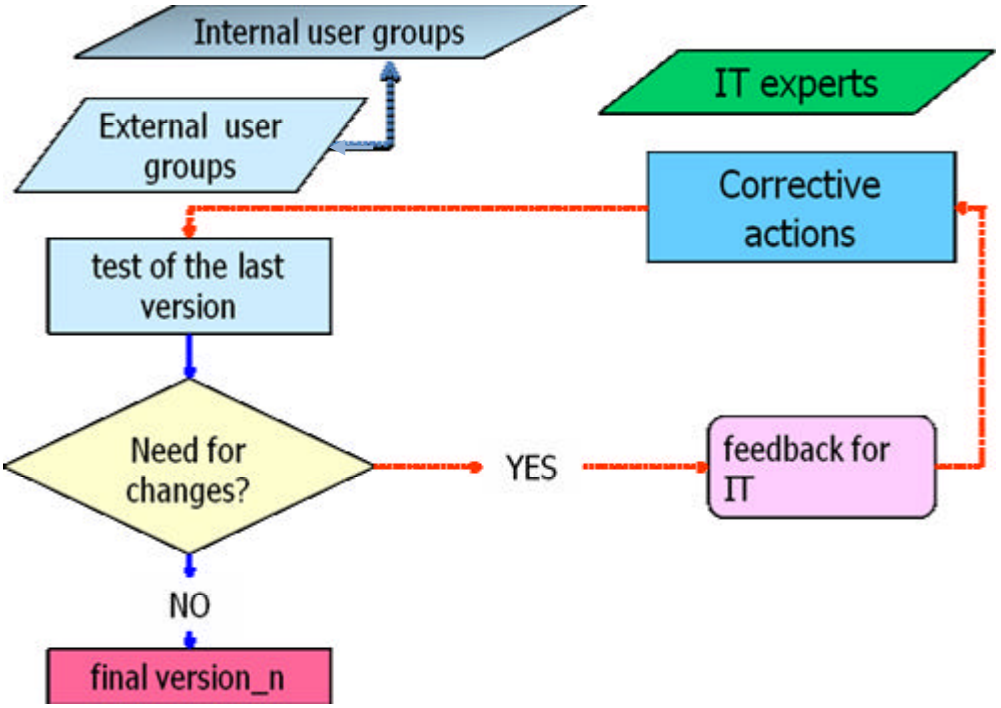


Figure 7. Second validation

2.2.3.2. Main System Functionalities

- Efficiency of application
- Effectiveness of application in supporting user-goals
- Technical robustness
- User satisfaction

2.2.3.3. Timeline of Validation

Activity	June 2008	July 2008	August 2008	September 2008	October 2008
1st prototype installation					
1st validation test					
Prototype modification					
2nd validation test					
Usable version					

2.2.3.4. Users Validation Groups

Two User groups involved in validating the MOSAICA system:

- "User Groups" within the project: in charge of the first validation and during the second validation they will act as appraisal managers and liaise with external user groups acting as the interface to the project.

Table 8. The profile of user groups within the project

No	Name	Affiliation	Profile
1	Clara Barrafranca	ORT	Senior project manager: Clara is in charge of training and pedagogical engineering in different European educational projects (R&D). She has been leading research and development in the areas of learning methodology and e-pedagogy. Clara has been leading pilot phases regarding different projects (from definition of criteria, to the implementation of the test, passing by the validation of the pilot, report results and corrective actions).
No	**Name**	**Affiliation**	**Profile**
2	Olivier Tassetti	ORT	Olivier Tassetti is IT manager for Ort France Paris since 2002. He is working in NTIC department. His job is to handle e-learning systems in Ort France schools; he also works in research and development of new learning systems for Education and Training European Projects, he is in charge of technical partners' coordination and provides advice and support to the project team.
3	Francois Wagner	ORT	Mr Wagner works as consultant and a trainer for ORT France**,** he studies the projects feasibility, conceive, carry out and parameterize softwares in scientific, technical and industrial data processing, install and follow up operating systems, train the users.
4	Elio Lumbroso	ORT	ICT project co-ordination, setting up and directing distance learning systems – consulting with pedagogical teams to introduce and develop ICT for E-Learning projects – co-operative work online between local ICT

			managers of ORT France.
5	Gabriel Elkaim	ORT	Historian - Geographer Person in charge of the training of the professors of Jewish history in ORT france. Assistant editor of studies at the Aquiba School (since 2006) Graduate of the Hebraic University of Jerusalem and the University of Letters and Social sciences of Strasbourg.
6	Raphael Attias		Raphael has taught for thirty years the Jewish history and the Jewish culture, Mathematics, business computer and new technologies of information and the communication. Aged of 57 years old, he was also director of the Moroccan Institute of the High Hebraic Studies in Rabat (Morocco).

- "User Groups" outside the Project : In charge of the second validation

Table 9. The profile of user groups outside the project

No	Name	Affiliation	Profile
1	Raphaël Rouckout	Consultant	Web designer – Integrator E-content designer. Master in Visual Communication Créapole-ESDI School management creation (design, communication, fashion, illustration, multimedia...).
2	Luis Manuel Pinto	CEJI	Luis Manuel Pinto is European *Educational Advisor* of the European Peer Training Organisation (EPTO) / Youth Initiative of the European Jewish Information Centre (CEJI) EPTO is a network of organisations and individuals who are working with peer

			education on diversity issues
3	Serge Freydman	ORT	Coordinator of teaching Hebrew to ORT; Professors Jewish culture and Hebrew to ORT. He has developed with other teachers two websites for the teaching of Hebrew and Jewish culture. He is precursory of various different pedagogical methods to enrich the Hebrew courses,(supports DVD, karaoke, MP3 …)

3. EVALUATION FRAMEWORK

Evaluation of advanced technologies pertains primarily to attributing a measure of the quality of user's satisfactory, and level of the projects' success. Evaluation can be distinguished between traditional, quantitative assessment on one side of the spectrum and authentic assessment, which involves a variety of qualitative approaches on the other side [12].

Tasks to be assessed should cover a broad spectrum of cognitive capabilities. Low level assignments might include retrieving information, recalling facts, and selecting an alternative from a list of possible options. Intermediate level tasks may call for exploring phenomena, manipulating equations, and analyzing the automatically collected results. High level requirements to be assessed may include drawing conclusions, transferring knowledge from one domain of discourse to another, posing relevant and complex questions regarding a subject under investigation, and build well structured and judgmental argumentations [12,13]. Our evaluation plan and tools, presented in the following sections were developed, keeping these various levels of assessment in mind.

The evaluation process assesses user opinions with regards to effectiveness and technological performances of the MOSAICA system achieving positive social awareness and cultural pluralism. Due to the service oriented structure of the system, it is important that the users evaluate the system in terms of each of the system components and specifically, technological performance of each of the specific components, and also evaluate these aspects with regards to the MOSAICA system as a whole.

The use of a collection of evaluation tools (online surveys, expedition diary, semi-structured interviews and access logs' analysis) is suitable to allow both generalized and specific evaluation of the system in terms of:

- user feedback on the applicability of the system to defined user groups
- use of core functionalities of the system from the user perspective
- assessing fulfilment of deliverables from the user perspective

- suggestions for future systems development
- user's disposition towards open-mindedness.

Data analysis and the preparation of reports due at the end of the project will be carried out based on quantitative and qualitative analysis considering all the users suggestions and comments. It is important to stress that the technological effectiveness can be considered a preparatory stage for a wider “trial” process starting after project final delivery and is essential for a system to become marketable and, above all, part of routine social practice.

Although this final phase requires a heavy and reliable user involvement (as only the comparison of the outputs from MOSAICA with the other similar tools, coupled with the professional socio-cultural and pedagogical experts who can validate an operational MOSAICA), the wider "trial" phase will not be part of MOSAICA project time frame. This is due to the fact that neither time nor resources within the MOSAICA project are sufficient for such a wider validation of the different components of MOSAICA. Technical validation of the different modules usability studies will provide all the inputs to write a wider validation plan.

As a base for our evaluation and assessment plan, we present a comprehensive theoretical framework focusing on evaluation of human attitudes, views and, behaviour. The theoretical underpinning presents a review of research methods for evaluation purposes, describes qualitative and quantitative research methods, focusing on the evaluation instruments used in this project.

Since one of MOSAICA's main goals is to develop a generic technological system for the preservation of cultural heritage, the literature review presents research on usability of web-based technologies, its importance to the user's experience, and ways to evaluate it. An additional main goal of MOSAICA is to present learning experiences that may enhance pluralistic views and open-mindedness among users. Therefore, we present literature review based on psychological and educational definitions of open-mindedness and ways to enhance and assess disposition towards it.

We start with a review on evaluation research methods, focusing on characterizing quantitative and qualitative assessment and the benefits of the mixed methods

approach. Next, we present a literature review on each of the evaluation tools, following a definition of usability and its importance in studying the usage of new technologies. We then argue that the ability to think critically and hold open-minded opinions is imperative for the attainment of a pluralistic society. Finally we present our plan for societal impact evaluation.

3.1 Evaluation Research Methods

The research process [14] usually starts with a broad area of interest, the initial problem that the researcher wishes to study. But, usually the initial interest is far too broad to study in any single research project; so the researcher has to narrow the question down to one that can reasonably be studied in a research project. This might involve formulating a hypothesis or a conjecture. (Deduction)

Alternatively data is collected and the researcher begins to try to understand it, usually by analyzing it in a variety of ways by attempting to address the original broad question of interest by generalizing from the results of this specific study to other related situations. (Induction)

We often refer to the two broad methods of reasoning respectively as the ***deductive*** and ***inductive*** approaches [figure 8].

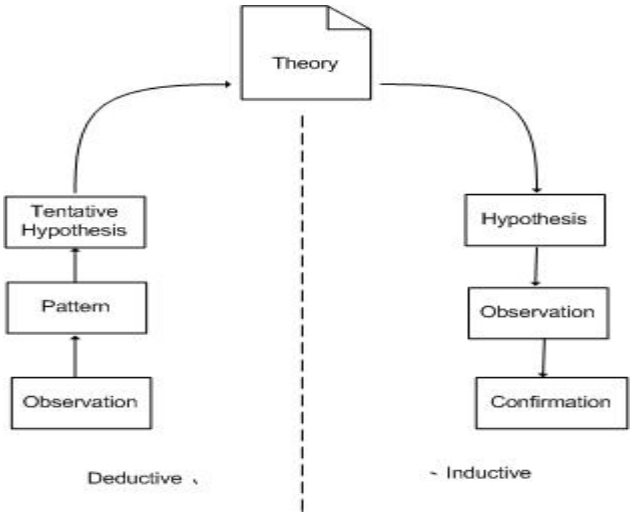


Figure 8: Deductive Vs Inductive Reasoning

Deductive reasoning works from the more general to the more specific. Sometimes

this is informally called a "top-down" approach. We might begin with thinking up a *theory* about our topic of interest. We then narrow that down into more specific *hypotheses* that we can test. We narrow down even further when we collect *observations* to address the hypotheses. This ultimately leads us to be able to test the hypotheses with specific data -- a *confirmation* (or not) of our original theories.

Inductive reasoning works the other way, moving from specific observations to broader generalizations and theories. Informally, we sometimes call this a "bottom up" approach. In inductive reasoning, we begin with specific observations and measures, begin to detect patterns and regularities, formulate some tentative hypotheses that we can explore, and finally end up developing some general conclusions or theories. Inductive reasoning, by its very nature, is more open-ended and exploratory, especially at the beginning. Deductive reasoning is narrower in nature and is concerned with testing or confirming hypotheses.

Evaluation may be defined as the systematic acquisition and assessment of information to provide useful feedback about some object. Evaluation utilizes many of the same methodologies used in traditional research methods, but because evaluation takes place within a socio-political and organizational context, it requires group skills, management ability, political dexterity, sensitivity to multiple stakeholders and other skills that social research in general does not rely on as much.

The generic goal of most evaluations is to provide "useful feedback" to a variety of audiences including sponsors, donors, client-groups, administrators, staff, and other relevant constituencies. Most often, feedback is perceived æs "useful" if it aids in decision-making and future improvements.

Evaluation strategy means broad, overarching perspectives on evaluation. They encompass the most general groups or "camps" of evaluators; although, at its best, evaluation work borrows eclectically from the perspectives of all these camps. There are multiple groups of evaluation strategies - ***qualitative/anthropological models***, that emphasize the importance of observation, the need to retain the phenomenological quality of the evaluation context, and the value of subjective human interpretation in the evaluation process. Included in this category are the approaches known in evaluation as naturalistic; the various qualitative schools; critical

theory and art criticism approaches; and, the 'grounded theory' approach of Glaser and Strauss among others. Another is ***participant-oriented models***. As the term suggests, they emphasize the central importance of the evaluation participants, especially clients and users of the program or technology. Client-centered and stakeholder approaches are examples of participant-oriented models, as are consumer-oriented evaluation systems.

There are many different *types of evaluations* depending on the object being evaluated and the purpose of the evaluation. Perhaps the most important basic distinction in evaluation types is that between ***formative*** and ***summative*** evaluation. Formative evaluations strengthen or improve the object being evaluated -- they help form it by examining the delivery of the program or technology, the quality of its implementation, and the assessment of the organizational context, personnel, procedures, inputs, and so on. Summative evaluations, in contrast, examine the effects or outcomes of some object -- they summarize it by describing what happens subsequent to delivery of the program or technology; assessing whether the object can be said to have caused the outcome; determining the overall impact of the causal factor beyond only the immediate target outcomes; and, estimating the relative costs associated with the object.

3.2 Quantitative and Qualitative Research and Tools

Researchers have long debated the relative value of qualitative and quantitative inquiry [15, 16]. Qualitative research, uses a naturalistic approach that seeks to understand phenomena in context-specific settings, any kind of research that produces findings not arrived at by means of statistical procedures or other means of quantification. Logical positivism, or quantitative research, uses experimental methods and quantitative measures to test hypothetical generalizations. Where quantitative researchers seek causal determination, prediction, and generalization of findings, qualitative researchers seek instead illumination, understanding, and extrapolation to similar situations [16].

Quantitative research methods were originally developed in the natural sciences to study natural phenomena. Examples of quantitative methods accepted in the social sciences include survey methods, laboratory experiments, formal methods (e.g.

econometrics) and numerical methods such as mathematical modeling. Quantitative researchers collect 'facts' and study one relationship of 'facts' with another. They measure using scientific techniques that are likely to produce quantified and general solutions. In quantitative research methods the test group normally consists of a large group of participants. The main purpose being to deliver representative results [16].

Exemplar quantitative research methods include questionnaires (online or otherwise) containing primarily open questions. The answers for these questions have already been compiled by the producer of the questionnaire. The task of the test subject is to find those answer(s) that best correspond to their opinion.

Qualitative research methods are developed in the social sciences to enable researchers to study social and cultural phenomena. Examples of qualitative methods are action research, case study research and ethnography. Qualitative data sources include observations, interviews and open-ended questionnaires, documents and texts, and the researcher's impressions and reactions. Qualitative research is focused on social process and requires the author to think abstractly and critically analyzing situations, obtaining valid and reliable information, and a sense of observation and interaction [17].

Qualitative research is common in social and behavioural sciences. It is often used to describe the culture of an organisation, system or community. Qualitative research provides an overview of the broad range of non-numerical measurement approaches. There are various qualitative **approaches** such as the *ethnographic approach* and the *Grounded theory*. The *ethnographic approach* to qualitative research comes largely from the field of anthropology. The emphasis is on studying an entire culture. Originally, the idea of a culture was tied to the notion of ethnicity and geographic location (e.g., the culture of the Trobriand Islands), but it has been broadened to include virtually any group or organization. That is, we can study the "culture" of a business or defined group (e.g., a Rotary club). *Grounded theory* is a qualitative research approach that was originally developed by Glaser and Strauss in the 1960s [18]. The self-defined purpose of grounded theory is to develop theory about phenomena of interest. As the researcher begins to gather data, *core theoretical concept(s)* are identified.

Although most researchers have traditionally advocated one approach or another, today, a growing number of researchers suggest combining qualitative and quantitative research methods into one study [15, 16, 19, 20, 21, ,]. Indeed Martin and Turner [21] claim that the two approaches are complementary to the extent that they should not be adopted in isolation.

Strauss and Corbin [16] claim that qualitative methods can be used to better understand any phenomenon about which little is yet known. They can also be used to gain new perspectives on issues about which much is already known, or to gain more in-depth information that may be difficult to convey quantitatively. Thus, qualitative methods are appropriate in situations where one needs to first identify the variables that might later be tested quantitatively, or where the researcher has determined that quantitative measures cannot adequately describe or interpret a situation.

The Mixed methods approach combines quantitative and qualitative data. The mixed methods research draws from the strengths of both quantitative and qualitative research, and strives to minimize the weaknesses of both in single research studies and across studies [22]

On the basis of combining qualitative and quantitative methods, we chose to use a variety of instruments in order to evaluate MOSAICA website and the VEs. The quantitative tools include closed-item questionnaires and the analysis of access logs. The qualitative tools include semi-structured interviews among users and the content analysis of the expedition diary. All tools aim at testing cognitive, affective, and operative aspects concerning MOSAICA and VE exploration. The analysis of each tool and the synthesis of all the findings together will produce an important overview of the usability of MOSAICA and the VEs, as well as portray possible changes in users attitudes towards more pluralistic and open-mindedness views.

3.2.1. *Online Questionnaires and Surveys*

The terms questionnaire and survey are often used interchangeably, however, if a distinction is to be made, a questionnaire is a form of questions people fill in, while a survey is a more general term that might include this as well as face to face or

telephone interviews. Unlike semi-structured interviewing, surveys follow a very specific and structured set of questions. Questionnaires and surveys will usually require some professional help to make sure that they are worded correctly and can be properly analyzed. They are a good way of gathering quantitative information but often reveal less about people's perceptions and concerns.

Structured questionnaires have the great advantage that a large number of stakeholders can be consulted and the cost of the realisation and analysis are low, especially through the use of closed questions. Furthermore, it is straightforward to draw comparisons and generalisations and the criteria of objectivity and reliability can approximately be fulfilled. The disadvantage of this method primarily lies in the fact that the interviewees are directed to answer the questions posed with the already given answers – the criterion of validity need not be assured. Using this method, the researcher who is going to evaluate the system must already know the system quite well. The primary effort lies in the work required to formulate the correct questions and choice of answers.

Technologically, evaluation is shifting from paper and pen to computerized questionnaires and surveys. In the recent years, with the increasing use of the Internet, online questionnaires have become a popular way of collecting data and information. Online tests offer convenience, efficiency, aesthetic and pedagogic improvements [23]. The role of information technology in evaluation and assessment has been growing rapidly [24] [25]. The high speed and large storage capacities of today's computers, coupled with their rapidly shrinking costs, makes online surveys a promising alternative to traditional paper-and-pencil measures.

Web-based testing systems offer the advantages of computer-based testing delivered over the Internet. The possibility of conducting an examination where time and place are not limited, however time and pace can still be controlled and measured, is one of the major advantages of web-based testing systems [26]. Other advantages include: the easy accessibility of on-line knowledge databases and the inclusion of rich multimedia, and interactive features such as color, sound, video, and simulations [23].

The design of an online questionnaire often has an affect on the quality of data gathered. There are several important paradigms that should be considered when

creating an online questionnaire, including the questionnaire format, its length, the type (open/closed) and the order of the questions, rating scales and the format in which the results will be presented.

Studies show that response rates are usually low and that there is a danger that they continue to drop due to over-surveying of web-users [27].

3.2.2 Online Knowledge Sharing Systems

Web 2.0 movement emphasizes the use of World Wide Web technology and web design for enhancing collaboration among users, especially, knowledge sharing. Web 2.0 movement led to the development and evolution of web-based communities and hosted services, such as social-networking sites, wikis, blogs, and folksonomies, that provides new forms of interaction via email, forums, group teleconferencing, and collaborative work.

In recent years, discussion on educational technology centers not only around instructional design and the learning management systems, but mostly on approaches that dramatically shift the centre of learning from teachers to students [28],[29][12], [30]. Social networking applications (blogs and forums), informal learning, e-portfolios, and personal learning environments (PLE) draw from Web 2.0 movement.

In the spirit of web 2.0, user-friendly blogs and discussion boards designed to encourage users' involvement in knowledge sharing and discourse, are becoming commonplace [31][32]. Online forums and blogs provide an important opportunity for users to present their thoughts and opinions [32]. MOSAICA's Expedition diaries are designed as web 2.0. applications generated for posting user's thoughts, comments and recommendations, as well as holding discussions among VE visitors, as an advanced form of online diaries.

[33] noted that active engagement in online environments emerges when users post their ideas online. The skills brought to the online forums are specific strategies that users can use to focus and move the discussion forward. Knowledgeable use of such strategies can enrich and deepen the dialog and foster better understanding of the emerging venue, which in MOSAICA's case is the preservation of cultural heritage.

Online communities may not only help in knowledge sharing and the creation of

communities of practice, but also to enhance potential leaders. Many capable people who lack natural charisma in live context may present a different opportunity online. Creating a virtual expedition and administrating a blog-based diary were users may also write their suggestions and comments makes the VE author a leader in a community.

3.2.3 Semi-Structured Interviews

Semi-structured interviews are part of the qualitative research method, used first in the social sciences, but today, used in fields such as education, law, business and medicine. Qualitative interviews may be used either as the primary strategy for data collection, or in conjunction with observations, document analysis, or other techniques [34].

Qualitative interviewing utilizes open-ended questions that allow for individual variations. An interview guide or "schedule" is a list of questions or general topics that the interviewer wants to explore during each interview. Although it is prepared to insure that basically the same information is obtained from each person, there are no predetermined responses, and in semi-structured interviews the interviewer is free to probe and explore within these predetermined inquiry areas [35]. Indeed, while a structured interview has a formalized, limited set of questions, a semi-structured interview is flexible, allowing new questions to be brought up during the interview as a result of what the interviewee says. Semi Structured Interviews are considered to be the most commonly used interview technique in qualitative research.

Online interviews are today a form of online research method that is growing rapidly. It takes many of the methodological issues raised in traditional interviews and transfers them online. Online interviews can be either synchronous interviews (chat technology) or asynchronous interviews (via email). They can be distinguished according to the number of interviewees that participate, as online interviews can be conducted on a one a one-to-one basis or in a group setting (online focus groups).

Whilst there is a vast body of literature concerned with qualitative interviewing, the online approach to interviewing remains a new and innovative research method (Volle, 2005)[36]. Web 2.0. innovative movement is based on the idea that the web is a place where people can create and communicate – in other words, to network. Related to

research and evaluation, Skype, for example, became a popular way to record online interviews and conversations. Indeed, there are many reasons why online interviews can be an appropriate and valuable methodological tool. Online interviews, as opposed to onsite interviews, provides the researcher with opportunities to interview individuals or groups who are often difficult to reach, such as the less physically mobile, geographically dispersed, living in dangerous places. It can also be used to reach a target audience where the audience is unknown, such as internet users and it provides savings in costs associated with human resources and travel.

3.2.4 Access Log Analysis

As more organizations view the Web as an integral part of their operations and external communications, interest in the measurement and evaluation of Website usage via log analysis is increasing. Compiled and interpreted properly, log information provides a baseline of statistics that indicate use levels and support use and/or growth comparisons among parts of a site or over time [37]. According to Anderson and colleagues [38], there are five main sources for analysis data: counters, cookies, user log-on forms, web-based site analysis services and server log files influenced by two important impacts on server log data: caching and IP addressing.

The shortcomings of log files as usage indicators fall into three main categories: certain types of usage data are not logged; the data that are logged may be incomplete; and it is tempting to draw unsound inferences from some of the data [37]. More sites and more people surfing the Internet create network congestion, making pages load more slowly. This is tempered somewhat by caching, a process which speeds up document loading time by automatically storing copies of requested files at different physical locations that are more readily accessed than the actual server. While caching speeds up user access to content, the process limits the validity of log files as an estimate of site traffic because once a file is cached (after a first visit), repeat visits are not recorded in the log file [38].

A "user", as reflected in a log, is an IP address—a computer. This does not necessarily correspond in a one-to-one ratio with an individual person. Therefore, a single IP address might actually represent several different users. And a single user might be associated with several different hosts if they access the Internet at different times or

from different computers or service providers (e.g., from work or from home). In addition, one user requesting a single page with one text element and nine graphic elements could be identified in the log file as 10 different computers accessing a single page [39]. Nevertheless, log file analysis, if performed carefully, can provide useful benchmarks and comparative measures to improve site structure, management, and usability.

3.3 Social Impact Evaluation

It is generally agreed upon that the ability to think critically is becoming an imperative to success in modern life, as the pace of change continues to accelerate, and complexity and interdependence continue to intensify. Nowadays people are not expected to 'know their place' but to determine and regenerate their own opinions and views [40]. As the world progresses, more and more people are required to make rational decisions based-on evaluative/critical thinking rather than to accept authority. One aspect of critical thinking is open-mindedness. A person who wishes to think critically about something like politics or religion should be open-minded to different views and cultures. This requires being open to the possibility that not only others are right, but also that you might wrong.

3.3.1 Open Mindedness

Rokeach (1960) defines open-mindedness as the extent to which the person can receive, evaluate, and act on relevant information received from the outside, on its own intrinsic merits, unencumbered by irrelevant factors in the situation arising from within the person or from the outside. Hare [41] offers open-mindedness as one of the eight qualities central to the educational enterprise: humility, courage, impartiality, open-mindedness, empathy, enthusiasm, judgment, and imagination. Although each of these eight qualities is important in its own right, each is required to be sometimes constrained, sometimes supplemented, and sometimes complemented by one or more of the other seven.

Open-mindedness requires that the individual would be prepared to entertain the thought that he might be wrong, and must be willing to revise his ideas in light of new information. Hare [41] argues for open-mindedness as an educational aim. According to him, open-mindedness is often thought to encourage credulity and to discourage

the formation of definite views. The aim of open-mindedness has been criticized by Gardner [42]. He has reservations about the educational desirability of encouraging children to be open-minded, indicating that teaching children to be open-minded leads to the discouragement of firm beliefs.

Hare and McLaughlin [43] dismiss Gardner's anxieties as unsubstantiated. They defend their philosophy of open-mindedness by arguing that it contingent upon the notion of appropriate firm belief. Some beliefs, Hare and McLaughlin argue, are on a continuum of beliefs and threaten the limits of open-mindedness. In addition, Hare and McLaughlin [43] emphasize that they do not believe that one can or should be open-minded about any and every belief that one holds. They claim that some beliefs justifiably inhibit dogmatic hesitancy and doubt. However, their conception of open-mindedness is strongly related to and presupposes the norms of rationality, which are integral to a disposition to revise beliefs in the light of evidence and argument.

Jonassen, Carr & Yuel [44] make a connection between open-mindedness and technology progress. They claim that technologies that have been developed by instructional designers are often marketed to educators as "validated" and "teacher proof", removing any meaningful control of the learning process by the learners or the teachers. They argue that technologies should not support learning by attempting to instruct the learners, but rather should be used as knowledge construction tools that students learn with, not from. In this way, learners function as designers, and the computers function as Mindtools for interpreting and organizing their personal knowledge.

Jonassen, Carr & Yuel [44] introduce the concept of Mindtools as computer applications that, when used by learners to represent what they know, necessarily engage them in critical thinking about the content they are studying. They provide several types of Mindtools such as: semantic organization tools, dynamic modelling tools, information interpretation tools, hypermedia and so on. Learners do not use Mindtools naturally and effortlessly. MOSAICA and its VEs are an example of a Mindtool. MOSAICA's applications not only present cultural heritage items, but requires users to participate in a quest, solve problems, make decisions, comment and give suggestions. While they are actively engaged with MOSAICA, users explore new

themes in an innovative way, thus enhance their higher-order thinking, and critical thinking. This process of learning may lead to a change in their opinions resulting in a more pluralistic view of the world, and the people and cultures in it. The pre-post VE questionnaires were designed and developed in order to examine changes in users' dispositions towards open-mindedness. Thus, examining whether a shift toward more pluralistic views occurred as result of using MOSAICA and exploring its VEs.

3.3.1.1 Development and Validation of Open-Mindedness Questionnaire

The Open-Mindedness Questionnaire is administrated before and after users explore MOSAICA applications, examining whether a shift in users' disposition towards open-mindedness took place. The questionnaires validation was conducted by four experts in educational research, from the Technion team. At first, 12 closed 1-to-6 scale items about open-mindedness inspired by CCTDI (The California Critical Thinking Disposition Inventory[5]) [45] open-mindedness scale were adopted. Through the validation process, some statements were reformulated to make them clearer to the users, and eight additional statements were added to meet MOSAICA's scope of activities and themes. In addition, to simplify the questionnaire, the scale was changed into a 1-to-5 Likert-type scale. The final 20 items were divided into 5 groups by their specification (4 items in each group): willingness to adopt different opinions, non-prejudice views, non-judgmental views, acceptance of others/strangers and gender equity. The questionnaire reliability was measured using Alpha Cronbach procedure, resulting 0.72 of internal consistency. The open-mindedness questionnaire is presented in appendix A.

3.3.2 Usability

Usability is a term used to measure how easily people can interact with a computer program in order to achieve a particular goal.

Usability refers to how people work with the product while testing usability means making sure that people can find work with the functions to meet their needs [46]. The way that users will work with the product is a basic issue for the product

[5] Facione, P.A. & Facione, N.C.(1992). The California Critical Thinking Disposition Inventory (CCTDI). Millbrae, CA: California Academic.

designers and developers.

The primary notion of usability is that the program we want to test should be designed with emphasis on three basic needs to be accessible to users:

Efficiency of use – the program should take less time to accomplish a particular task;

Easy to learn – operating the program should be easy to learn, by reading the instructions and observing the program;

Satisfying to use – using the program needs to be more fun and interesting than using other objects with the same purpose.

Complex computer systems are finding their way into everyday life. This has led to usability becoming more popular and widely recognized in recent years as companies see the benefits of researching and developing their products with user-oriented instead of technology-oriented methods. By understanding and researching the interaction between the user and the program, it is possible to design the program to fit better to the target users. By testing the program and observing potential users operate the program, one can identify needed functionality or design flaws that were not anticipated.

While there can be wide variation in where and how to conduct a usability test, every usability test shares five characteristics [46]:

The primary Goal is to improve the usability of the product.

The participants represent real users.

The participants do real tasks.

It is necessary to observe and record what participants do and say.

The data is analyzed and recommendations are raised for changes.

Usability testing generally involves measuring how well test subjects respond in four areas:

Time on Task - How long does it take participants to complete basic tasks?

Accuracy - How many mistakes did the participants make? (And were they fatal or recoverable with the right information?)

Recall - How much do participants remember afterwards or after periods of non-use?

Emotional Response - How does the participant feel about the tasks completed? (Confident? Stressed? Would the user recommend this system to a friend?)

The results of a first usability test can be treated as a baseline or control measurement; all subsequent tests can then be compared to the baseline to indicate improvement.

The data collected from the usability tests can be both quantitative and qualitative [47]. Quantitative data can be a measurement of accuracy, speed, and recall, while qualitative data can be the researcher's subjective descriptions of what is going on or the subjects own opinions. The qualitative data can be collected through both multiple-choice and open-ended questions.

Mayhew [48] argued that testing should focus on ease of learning as well as ease for use. In each case, a formal usability testing is given to three to ten representative users, running through a set of core, high-frequency, realistic test tasks. To test for ease of learning, minimal instruction should be given. To test for ease for use, adequate training and practice should be provided to simulate expert users.

According to Rubin [49], usability testing identifies 80% of the problems with a small group of people (up to eight representative users). It is not a formal research. The method is to spend time with the user and allow people to experience.

Nielsen and Landauer [50] claimed that five users are enough for running usability tests at various stages of the development process. Their argument was that the best results come from testing no more than five users and running as many small tests as one can afford. They showed, using a mathematical model that most usability problems are found in a usability test with five users. Most researchers today agree that, although five users can generate a significant amount of data at any given point in the development cycle, in many applications a sample size larger than five is required to detect a satisfying amount of usability problems.

We intend to adopt Nielsen and Landauer's approach, and we will focus, as a first stage, on five users for running a pilot usability study for examining the VE application**.**

3.3.2.1 Development and Validation of Usability and Learning Outcomes Questionnaire

Usability and Learning Outcomes Questionnaire is a post questionnaire, allowing users to rank and comment about cognitive (learning), affective (satisfaction) and operative (technical) issues related to MOSAICA usage. The items in this questionnaire were based on a research tool developed by Barak & Rafaeli [23] consisting of 12 statements on a 1-to-5 Likert type scale (1- strongly disagree, 5- strongly agree). The usability and learning outcomes questionnaire validation was based on researcher validation conducted by four experts in educational research from the Technion team. The questionnaire reliability was measured using Alpha Cronbach procedure, resulting 0.87 of internal consistency. The usability and learning outcomes questionnaire is presented in appendix B.

3.3.2.2 Development and Validation of Semi-Structured Interview Questions

In order to achieve a holistic evaluation of MOSAICA's applications and to conduct a more personal dialogue with the users, semi-structured interview questions were developed. The questions and their sequence are partially pre-determined, allowing new questions to be brought up during the interview as a result of what the interviewee says, thus leading to a more revealing and in depth interview.

The interview questions were based on the four principals of Dumas and Redish [46] usability testing:

The time spent in exploring MOSAICA, measured through observation.

Accuracy is observed while the users explore the VEs.

Recall is measured by several questions focused on content comprehension and higher order thinking skills (aspect 2 in the questionnaire, appendix C).

Emotional response is measured through questions focusing on the users' interest, enjoyment and personal opinions (aspect 1 and aspect 3 in the questionnaire, appendix C).

The interview questionnaire consists of three aspects: a. usage characteristics of the VE, b. content comprehension and higher order thinking skills and, c. open

mindedness and tolerance attitudes. The semi-structured interview questions are presented in appendix C.

3.3.3 Expedition Diary

The Expedition Diary tool aim is twofold. First, to enable the Virtual Expedition creator to provide commentary on the VE s/he created. Second, to allow other users to communicate with the VE creator and with fellow users. Thus, expressing their knowledge, thoughts, and feelings related to the specific VE theme, while creating a community of users that share knowledge.

3.3.3.1 Expedition Diary Structure

Each VE creator will have an option to write her remarks related to the VE she created and the VE users will be able to leave their comments after and during VE exploration.

Each Expedition diary will be maintained by the VE creator, with regular entries of commentary, descriptions of events, or other material such as graphics or video. Entries will be displayed in a reverse chronological order (last shows first). The VE will function as an online diary of a community of interest, providing commentary or news on a particular subject- the VE theme.

Each Diary will be defined by two keys – Virtual Expedition key and creator username. This approach will allow presenting all the comments per specific VE creator. Users will be encouraged to provide contact information for further correspondence.

3.3.3.2 Expedition Diary Design

The VE Diary will be a default part of VE creation process so that for each Virtual Expedition, the Diary will be created automatically (the VE creator has the option not to open an expedition diary). During the VE creation, the creator will be able to write remarks so that they will be visible to the whole MOSAICA community after Editorial Board approval of VE.

After the VE creation will be finished, the VE creator will still be able to add/change/delete her remarks. After Editorial Board approval, the VE users will be able to explore the published Virtual Expedition and to leave their comments.

The Diary will be opened by clicking on the "VE Diary" button available on each one of

the screens during VE exploration. In addition, presenting the VE as the result of MOSAICA ontology search will also allow opening the relevant VE Diary.

Expedition Diary tool will consist of three buttons:

Add a diary to your VE (for the VE creator)

VE Diary (for the VE users)

Links to other related VEs

"Add a Diary to your VE" will be the part where the VE creator will write her thoughts and ideas about the Virtual Expedition. Only VE creator will be able to modify this part. In order to modify the Expedition Diary, the text editor screen will be opened with ability to change between standard textual representation and HTML representation. The VE creator will be able to write text and add hyperlinks to audio, video and other file types.

There will be a possibility for customizable user interface. Each one of the Expedition Diary creators will be able to choose one of the graphical styles available in the MOSAICA database. Each graphical style will define the size and style of the font, the background image, the colors and all other parameters related to the information representation (TBD).

"VE Diary" will be the part where VE users will be able to share text messages, pictures, video, and sound files related to the presented VE theme. They will also be able to rank the VE on a scale of 1-to-5 and share their comments and suggestions (similar to the YouTube technology).

"Links to other VEs" will include links to other Expedition Diaries. It will consist of several frames:

Other Expedition Diaries of the same VE creator

The best Expedition Diary (according to the grade given by the VE users in the Comments part)

The most popular Expedition Diary (according to the number of comments)

3.3.4 *A summary of the social Impact evaluation tools*

As part of MOSAICA's evaluation and assessment plan, quantitative and qualitative methods are employed in the analysis and interpretation of data. To summarize, the evaluation will include the following tools:

An online survey – a closed item questionnaire, on a 1-to-5 like type scale that will be opened/accessed at the end of each Virtual Expedition. The survey will capture the visitor's views on and attitudes toward: (1) extent of familiarity with the Jewish culture and heritage, (2) multicultural society and tolerance, and (3) learning experience and the development of critical thinking.

Expedition diary – an open input prompt that will also be posted at the end of each Virtual Expedition, in which visitors can write their impressions, ideas, and remarks. A visitor will also be able to provide contact information for further correspondence. Content analysis of the visitors' narrations will allow the examination of their opinions and perceptions about MOSAICA, multicultural society, open-mindedness, and tolerance.

Semi-structured interviews – a dialogue with the visitors who will indicate willingness, and provide their contact details. The questions and their sequence in these interviews will be partially predetermined. The interviews will examine the way users apply higher order thinking, and construct meaning about Jewish heritage. The analysis will determine how visitors construct mental models while engaging in discovery and exploration, and whether and how their beliefs about a pluralistic society and tolerance are gradually reinforced as they engage in MOSAICA.

Log analysis – an objective tool for tracking the way visitors use MOSAICA. By accessing the logs we will be able to rank various resources via visitors' preferences, and study the patterns of usage in MOSAICA.

Assessment is typically divided into two types: summative and formative. MOSAICA's educational goals are assessed by both means. The formative assessment is used as a diagnostic tool to identify and improve areas of weakness. It is aligned with the constructivist theoretical framework, and associated with interviews and writing samples. The summative assessment will serve as a basis for overall judgment of the

project, relative to its objectives. Table 10 presents a summary of the way the evaluation tools are to be integrated as part of MOSAICA platform.

Table 10. Evaluation tools to be integrated as part of MOSAICA platform

Evaluation Type	Evaluation Tool	Goal	Description	Button	Action
Summative	Open-mindedness questionnaire	Evaluating users' dispositions toward open mindedness and pluralistic views, Before and after VE usage	a 1-to-5 online Likert type scale	Click here before login or logout A "blinking" button will draw the users' attention as he logs in or out of MOSAICA	Once clicking on the blinking button, the questionnaire opens in a separate window. The users' answers are collected on an excel sheet that can be opened and analyzed by the researchers.
Formative and Summative	Usability and learning outcome questionnaire	Evaluating users' experience using MOSAICA applications	a 1-to-5 online Likert type scale	Click here before logout A "blinking" button will draw the users' attention before he logs out MOSAICA	Once clicking on the button, the usability questionnaire opens as a separate window. The users' answers are collected on an excel sheet that can be opened and analyzed by the researchers.
Formative	Expedition diary	An open input tool, allowing communication between the VE creator and VE users forming a	Similar to You Tub technology	Add a diary to your VE (for the VE creator) VE Diary (for the VE users) Links to other	Once clicking on the add button, the VE creator is enabled to provide commentary on the VE s/he created. Once clicking on the VE

Evaluation Type	Evaluation Tool	Goal	Description	Button	Action
		"community of interest".		related VEs	diary button, the users are able to communicate with the VE creator and with fellow users.
Formative	Semi-structured interviews	An interview with MOSAICA users who will indicate willingness, and provide their Skype address.	Using Skype technology (or similar), allowing researchers to ask users evaluation questions.	Open Skype	The questions and their sequence in the online interviews will be partially predetermined. The analysis will determine whether and how users' beliefs about pluralistic society.
Summative	Log analysis	An objective way for tracking the visitors' use of MOSAICA.	Generating web pages for statistics of: hits; visits; countries; and more.	_	Indicating users' preferences and study patterns of usage. The statistics will be viewed graphically.

Indeed, MOSAICA is envisioned as a technological showcase that selected the domain of the cultural heritage for the demonstration purposes. To this purpose, we provide an independent, impartial evaluation of MOSAICA's technologies, and their potential societal impact when applied to other traditions and heritages, as presented in Figure 9 that summarizes the evaluation tools in a graphical way.

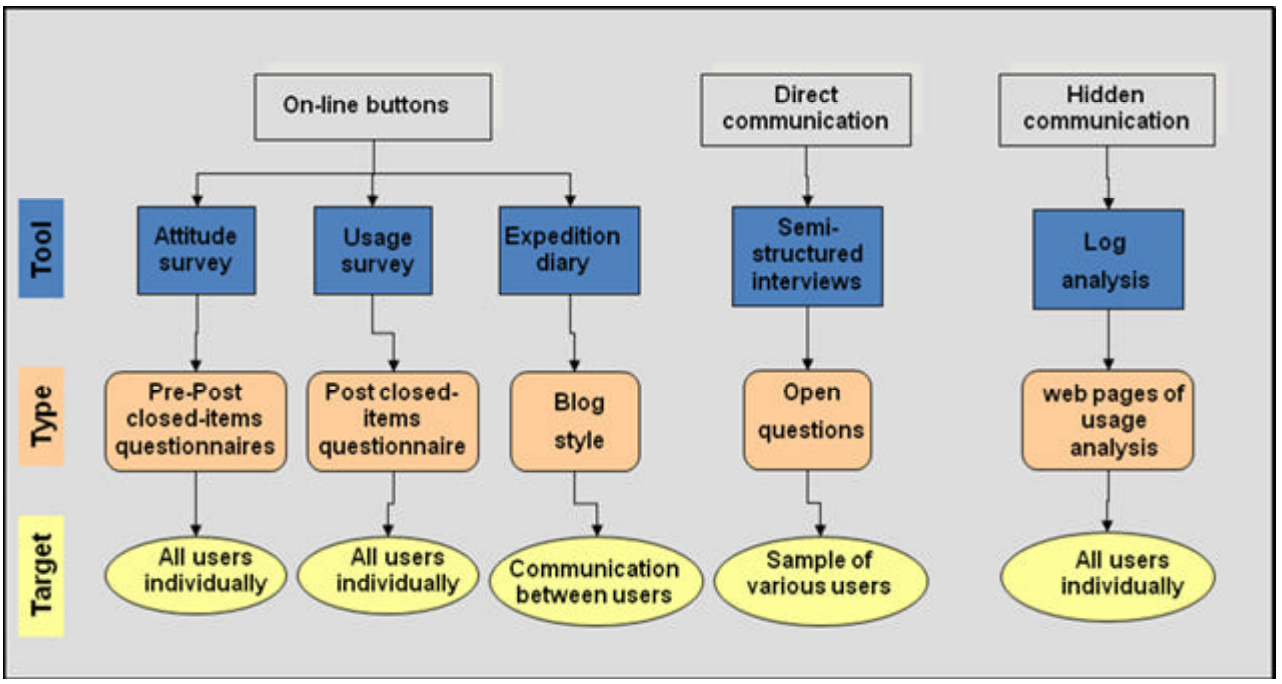


Figure 9. MOSAICA social impact, usability and learning outcomes evaluation

3.3.5 The Logistical Plan for Societal Impact Evaluation

For the purpose of societal impact evaluation, user groups including students in high schools and academic institutions, and museum visitors and employees, will be invited to a presentation that will be provided in conjunction with hands on exploration of MOSAICA. The meeting will be conducted as a two-hour workshop, providing an overview of MOSAICA system, allowing users to gain practical experience and to provide feedback related to the system's functionality.

Each workshop will include six parts:

(1) An expert presentation of MOSAICA system and its' functionalities (15 minutes).

(2)_Users respond to an online pre-questionnaire on disposition towards open-mindedness and pluralistic views (10 minutes).

(3) Users exploration of MOSAICA, including the semantic directory, GIS maps, and Virtual Expeditions, (40 minutes).

(4) Users write their experience, comments, and suggestions in an online "Expedition diary" (15 minutes).

(5) Users respond to an online post questionnaire on disposition towards open-mindedness and MOSAICA usability (15 minutes).

(6) Experts and users participate in a concluding discussion on MOSAICA system and its' usage (15 minutes).

The workshop will depict the different aspects of MOSAICA system that require user evaluation: functionality, usability, user satisfaction, and educational gains.

Participants are invited to use the MOSAICA system with demonstration of interfaces and utilities. Participants are provided with step by step instructions showing MOSAICA's explorative, guided and collaborative usages. All participants will be asked to answer the open-mindedness and usability questionnaires. In addition, a random sample of users (around 10% of each group) will be selected for responding to the semi-structured interviews.

3.4 Multi-Cultural Evaluation

3.4.1 Understanding Cross-cultural and Multi-cultural aspects of Evaluation [13]

When conducting evaluations in any community or organizational context, it is critical that evaluators understand what it means to be culturally aware and sensitive and to have some level of cultural competence. The first step toward achieving this goal is for evaluators to know themselves—their roots, histories, biases, prejudices, and assumptions about race, culture, and ethnicity. Only then can they begin to understand, and possibly confront, long-held beliefs that may support or impede their working within a multicultural environment. Activities carried out under a multicultural assessment framework include:

- Understanding the importance of recognizing the cultural contexts of evaluation
- Defining cultural competency within an evaluation context
- Understanding how one's culture and history potentially affect evaluation practice

3.4.1.1 Definitions

The following definitions are useful when discussing different aspects of culture with regard to evaluation practice.

Cultural Knowledge - Familiarization with selected cultural characteristics, history,

values, belief systems, and behaviours of the members of another ethnic group [51].

Cultural Awareness - Developing sensitivity and understanding of another ethnic group and involving changes in attitudes and values. It also reflects an openness and flexibility in working with others of another culture [51].

Cultural Sensitivity - Knowing that there are differences and similarities among cultures without making value judgments of good or bad, better or worse, right or wrong [52].

Cultural Competence - A set of congruent behaviors, attitudes, and policies that come together in a system, agency, or among professionals that enables that system, agency, or those professionals to work effectively in cross cultural situations [18]. Cultural competency is the ability to effectively operate in different cultural contexts.

3.4.1.2 Understanding the Influence of Culture in Evaluation Practice

This activity asks participants to consider how their own cultural heritage and background may influence the ways in which they design and implement evaluation studies. The objectives of this activity are to enable participants to

- identify or recognize their own cultural heritage or background
- understand other individuals' hopes, dreams, disappointments, and tragedies as they relate to one's culture
- consider how one's background may influence his or her evaluation practice

Under this activity participants are generally asked to briefly describe an object/artefact/incident/event that reflects something of their ancestry or cultural/ background and racial belonging. Storytelling method is widely accepted for such activity.

3.4.1.3 Defining Cultural Competence within an Evaluation Context

This activity asks participants to consider various definitions of evaluation how they relate to working with culturally diverse audiences. At first, decisions must be taken on various definition of evaluation within a cultural context. Several types of

definition are adopted in practice.

Commonly Used Definition - Program evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming [53].

World Bank Definition - Monitoring is defined as "the continuous assessment of project implementation in relation to agreed schedules and of the use of inputs, infrastructure, and services by project beneficiaries". *Evaluation* is defined as "periodic assessment of the relevance, performance, efficiency, and impact (both expected and unexpected) of the project in relation to stated objectives"[54].

Developmental Evaluation - Developmental evaluation consists of evaluation processes and activities that support program, project, product, personnel, or organizational development, usually all five. The evaluator is part of a team whose members collaborate to conceptualize, design, and test new approaches in a long-term, ongoing process of improvement, adaptation, and intentional change. The evaluator's primary function on the team is to elucidate team discussions with evaluative data and logic and to facilitate data-based decision making in the developmental process [55].

Inclusive Evaluation - Inclusive evaluation involves a systematic investigation of the merit or worth of a program or system for the purpose of reducing uncertainty in decision making and for facilitating positive social change for the least advantaged. Thus, inclusive evaluation is data based, but the data are generated from an inclusive list of stakeholders, with special efforts to include those who have been traditionally underrepresented. It does not exclude those who have been traditionally included in evaluations [56].

Feminist Evaluation -Feminist evaluation includes judgments of merit and worth, application of social science methods to determine effectiveness, and achievement of program goals as well as tools related to social justice for the oppressed, especially, although not exclusively, for women. Its central focus is on gender inequities that lead to social injustice. It uses a collaborative, inclusive process and captures multiple perspectives to bring about social change [57].

Basically, this activity requires participants to consider various definitions of evaluation as mentioned above and decide which sort of evaluation they want to apply in the context of evaluating MOSAICA's on culturally diverse audiences. To that aim participants will

- describe the distinguishing characteristics of various definitions of evaluation
- determine the ways in which these evaluation definitions relate to power
- discuss the importance of the definitions in relation to working in culturally diverse communities and the advantages and disadvantages of using the various definitions

Participants will be asked to form *small groups* of 3 to 5 persons each. Each of groups will be asked to read and discuss the definitions among its members and to come up with their decision based on the following:

- What are the implications of the various definitions for working in culturally diverse communities?
- What are the advantages and disadvantages of adopting each definition in a culturally complex setting?
- To what extent, or in what ways, has your definition of evaluation shifted or changed?

Then each group will be asked to share their comments and discuss the following issues to come up with their recommendation of definition that will be finally adopted-

[1] What is the essence of each definition?

[2] What similarities and differences do you see among the definitions? What distinguishes one definition from the other?

[3] What kinds of power rest in these definitions, and how might such definitions affect the design and implementation of an evaluation?

3.4.1.4 Cultural Sensitivity in Evaluation

This activity asks participants to consider how various aspects of culture affect an evaluation's design and implementation. In particular this activity focuses on two

things. Firstly, *cultural influences* that requires attention to cultural factors that might affect evaluation's design and implementation (e.g. communication styles, values and norms, learning preferences etc.) as well assurance of applying appropriate culturally sensitive data collection activities (observations, interviews and surveys). Secondly *Strategies for Making Evaluation More Culturally Sensitive* requires ensuring the following -

- Assess your own attitudes, beliefs, and values; be aware of your own biases.
- During the initial stages of the evaluation, explore any cultural issues that could affect the evaluation's design and implementation.
- Learn all you can about the cultural group(s) involved in the evaluation.
- Incorporate culturally diverse groups and perspectives throughout the evaluation process (e.g., obtain feedback on the evaluation's design, methods, and draft reports).
- Recognize that the evaluation may take more time to conduct if you are negotiating access and building trust in working with community groups and organizations.
- Use as little jargon as possible.
- Include representatives of culturally diverse groups on evaluation advisory committees.
- Be flexible in your choices of methods and activities.
- Use multiple data collection sources and methods.

Specifically the objective of this activity is to see how participants

[1] Reflect on their own values, assumptions, and experiences as they relate to working with different cultures

[2] Develop greater awareness of how ignoring the topic of cultural differences may affect an evaluation's design and implementation

[3] Determine ways in which various data collection methods can reflect cultural considerations

Participants will be asked to work in small groups (e.g. 3-5 persons) and discuss their responses to *Cultural Influences* and *Strategies for Making Evaluation More Culturally Sensitive.* Finally, assessment can be done based on the following questions -

– What else do you need to do to build cultural sensitivity in your evaluation work?

– What are the implications of not being culturally aware, sensitive, or competent?

– How do you think evaluators can develop these competencies?

3.4.1.5 Defining Cultural Competence: Evaluation in Multicultural Settings

This activity asks participants to consider the concept of cultural competence and how it relates to evaluation practice in the context of MOSAICA.

Within an evaluation context, the concept of culture includes shared behaviours, values, beliefs, attitudes, and languages and is broadly construed to be multidimensional in nature, based on such characteristics as race/ethnicity, gender, disability, economic class, and sexual orientation. Cultural competence might be expressed in different ways. *Cultural competence* refers to an ability to provide services that are perceived as legitimate for problems experienced by culturally diverse populations and this definition [58] denotes the ability to transform knowledge and cultural awareness into interventions that support and sustain healthy participant system functioning within the appropriate cultural context. Alternatively [59] cultural competence may be defined as a set of congruent behaviours, attitudes, and policies that come together in a system, agency, or among professionals and enables that system, agency, or those professionals to work effectively in cross cultural situations. The root of cultural competency in evaluation is a genuine respect for communities being studied and openness to seek depth in understanding different cultural contexts, practices, and paradigms of thinking. This includes being creative and flexible to capture different cultural contexts, and a heightened awareness of power differentials that exist in an evaluation context. Important skills include the ability to build rapport across difference, gain the trust of community members, and self reflect and recognize one's own biases.

The aim of this activity is to enable participants to

- Understand various definitions of cultural competency
- Determine why it is important for evaluators to consider definitions of cultural competence
- Understand the relationship between cultural competence and issues of social justice in evaluation

Participants will be asked to work in small groups with a minimum of 3 to persons in each group to discuss on how we should construct a Definition of Cultural Competence in the Evaluation Context of MOSAICA and groups will then share and highlight their thoughts that may be assessed based on the following factors -

[1] What does cultural competency look like? In other words, what would you see evaluators doing that exemplified cultural competence?

[2] What will you do to increase your own cultural competency?

3.4.1.6 Exploring the Role of Assumptions in Evaluation Practice

This activity provides an opportunity for participants to think about the assumptions we make about other people based on how they look or act and how these might influence the ways in which we design and implement evaluations. The main objectives of this activity is to enable participants to

- Reflect on how their own assumptions influence the design and implementation of an evaluation
- Discuss the ways in which their own assumptions can be tested and corrected within an evaluation context

This activity may be conducted in either in small groups or individual basis and assessment might be based on the following factors -

[1] What kinds of assumptions might we make about evaluation of features, functionality and impact of MOSAICA?

[2] What are the implications of conducting an evaluation with a certain set of assumptions about a group of people, a program, or a geographic area?

[3] How can evaluators learn to test and correct these assumptions in their practice?

3.4.2 Guiding Principles

To deepen its efforts to explore the intersection of MOSAICA initiative and cultural pluralism, we adopt the following guiding principles:

- Inclusion in design and implementation
- Appropriate Measures and Success
- Cultural and system analysis

3.4.2.1 Inclusion in Design and Implementation

Under this principle we have engaged the project partners related to MOSAICA content selection (WP1), design and implementation of sub-components and entire systems (WP2, WP3, WP4, WP5) to convene a large and diverse, intergenerational group of community leaders, museums, archives, and practitioners in the fields of community content archiving, organizing, community development, and pedagogical institutions - traditional and virtual. The three layer multifunctional MOSAICA systems enables semantic search and browsing within the cultural heritage domain in the first tier, virtual expeditions for pedagogical uses in the second tier and collaboratively upload and share cultural artefacts in the third tier. The **early design inputs** and functional requirements have been derived from user requirements based on the formulation and study of use cases, leading to the formulation of application scenarios.

The design and system requirements analysis team outlined following **specific recommendations** for achieving project's ultimate objective for promoting cultural pluralism and multicultural spirit:

- MOSAICA System must be *flexible* enough to allow the installation of different types of heritage items, that is, visual representations (photo, video, scanned documents, pictures etc.) with their adjunct text
- Users' privacy protection and choice of sharing material within a community must be handled within the system with a screening mechanism and allow

authorized users to evaluate the uploaded items and configure the sharing level of each item.

- Attracting users through MOSAICA's easy to use interfaces for search, browsing and virtual expeditions (VE) on GIS/TIS maps (including games, and inquiry assignments) and VE maker (uploading, assessing and validating Heritage items)
- The MOSAICA's exploitation has been aimed for multi-community users in different institutions such as school, museum, archives etc. As a result, the system must enable management of certain group of users by independent administrators in a particular community by generating users' templates, adding new groups and managing their contents.

The inclusion of above mentioned recommendations will be evaluated through measuring the MOSAICA applications' and created interfaces traceability to the requirements as stated in Deliverable D5.1- MOSAICA System and Application Specification and D4.1- MOSAICA Virtual Expedition Methodology.

Involvement of key stakeholders - Firstly, based on their role during the design phase, designers and users (as **evaluators**) from diverse communities will be considered carefully during the evaluation process. Secondly, appropriate **survey tool** and questionnaires will be devised to assess the user behaviour system usability particularly for the multicultural aspects. Thirdly, we will continue ongoing efforts to improve the evaluation tools, methods, and procedures. Finally, we will share evaluation results in different formats with different community of users throughout the evaluation, soliciting feedback from them while also helping them to interpret the correspondences between MOSAICA interfaces and the assessment tools. For instance, we will provide each evaluator (individual or a group) with individualized survey tools and guide them through a process of using the application interface and their need of data to give a feedback as required by assessment tools. We will also present them with tutorials of using the assessment tools on several different occasions.

Involvement of multi-community users in the evaluation - Our overall approach is to carry out detailed evaluation through interviews and questionnaires

involving two types of users - "small" group of naive users and "chosen expert" users from different faiths and cultures including Christians, Muslims, Hindus, Sikhs etc and those of no faiths at all. WE intend to do at least two evaluations of one pilot inviting participant from few organizations to help us with data collection and interpretation. In addition to hiring users, we will train interviewers to record interviews using the provided template. We will also engage people at several sites as "ethnographers," asking them to interview their peers.

Implementation of culturally appropriate evaluation strategies - We were intentional in using multicultural evaluation approaches to understand diverse organization and their user model of different cultural community. Internally, we will involve few highly diverse researchers in data collection and analysis, and to the extent possible, matched team members' expertise, background, and experience to the MOSAICA target user model, which worked with Asian, African, Latino and other cultural groups. We will also collaborate with a Blackfoot American researcher to help us to modify our evaluation design by visiting the MOSAICA portal site. These researchers will provide an external check on data collection and findings, and provide comments on our protocols, reviewed, and contributed to MOSAICA site profile. Validity stems from the variety of cultural perspectives and interpretations that will be solicited in team meetings and review by targeted user evaluation communities.

3.4.2.2 Appropriate Measures and Success

The MOSAICA portal and its interfaces will be assessed on the basis of the following:

- How MOSAICA application will provide an effective approach for reaching cultural pluralism among people and communities' adherent to different cultures not engaged collaborative access to cultural resources and artifacts?
- How does MOSAICA contribute to positive multicultural development in the society, including identity development & engagement in positive social exchange and tolerance in civic life?

The need for building the design and frameworks for evaluation lies in the importance of being explicit about the assumptions & philosophies as well as importance of community input *and* diversity theories and research. Framework based on the works

of Paolo Friere's & Erik Erickson [60, 61] suggests racial identity theory that

- Builds on tradition of popular education and education promoting tolerance towards cultural diversity.
- Engages vulnerable or marginalized communities in ongoing "praxis"– or a process of reflection and action.
- Has a dual focus on individual and community change.

Evolving assessment framework will judge MOSAICA outcomes against the benefit from the input of chosen evaluators focusing on the following factors in terms of how the use of MOSAICA portal promotes multi-cultrism and solves barriers to achieve a tolerant society:

- *Coalition Building* - Does MOSAICA develop connections with people of different ethnic/racial groups?
- *Civic Challenges* - How MOSAICA's impact may help in reducing issues related to racial intolerance in a multicultural community?
- *Visual Social Changes* - Is it possible to derive a mapping between MOSAICA activities and activities that people in a particular community think important?

Thus we plan to develop measures based on experiences of community groups and conduct follow up study to assess the validity of the measures.

3.4.2.3 Cultural and System Analysis

One of the aims of MOSAICA evaluation is to assess the MOSAICA system developed initially considering Jewish Culture as a use case for continuous future improvement that will generate scope for its further enhancement to fit the system for any cultural context. Backed by such motivation we will conduct the following activities:

- Exploration of Issues - Explicit discussions of issues of racism and intolerances highlighting issues of race, class, gender, and sexual identity
- Peoples Voice on Choice - Discussions within the context of how knowledge gained from MOSAICA base may improve users everyday lives?
- Collective Processes and Leadership - Through workshops and tutorials we

would like to see how users may connect their personal experiences from MOSAICA interfaces to abstract concepts (i.e. racial identity, commonality and differences among multiple cultural heritages etc.)

- Use and Examination of Popular Culture of different communities - We will pursue further to identify issues that may arise for MOSAICA's influence on popular culture of different communities.

3.4.3 A Holistic Multicultural Evaluation Framework for MOSAICA

Cowles [62] mentioned an enhanced multi-cultural evaluation methodology that we will adopt as a framework for MOSAICA multicultural evaluation.

***Self-reflection and exploring own identity of evaluators*:** Everyone has a multicultural identity. Reflect on how your identity has developed and evolved over time. An authentic racial identity is a critical component to becoming culturally competent.

***Inviting critique from colleagues*:** It's difficult to accept critique and even more difficult to publicly solicit and acknowledge feedback. Sharing with peers enables us to be an objective, fair, and judicious evaluator.

***Examining the relationship between intent and impact*:** It is easier to focus on our intent regardless of the impact, but unintentional outcomes can be very much damaging disregarding its intent. There may be a mismatch between intent and the actual impact of any activity.

***Rejecting the myth of color blindness*:** Everyone is touched, moved, or affected by race the moment they enter the world. Race shapes how others see us and how we view ourselves.

***Recognizing that stakeholders' identity group*:** evaluators may have little control over how people experience or interpret their intentions. Stakeholders base their reactions on a lifetime of experiences (i.e., their own identity development). Try not to take stakeholders' reactions personally, but be aware that they may have an impact on the evaluation process.

***Building coalitions with evaluators of diversified background*:** Consciously

developing multicultural coalitions, even in informal settings with evaluators who differ in race, ethnicity, sexual orientation, gender, religion, first language, disability, and other identity facets affect evaluation activity positively.

***Acknowledging the role of a social activist*:** Social activist as evaluators help determine the merit, worth, and effectiveness of programs and policies that affect multicultural populations.

***Educating own self*:** Educating oneself with adept knowledge about the diverse social and cultural groups that compose society is fundamental to being affirmative and inclusive. Promotion of pluralistic attitude may be achieved by exploring different people on multiple dimensions and in different places.

***Taking responsibility*:** Taking accountability for role played by evaluators in interactions during the evaluation process improves the conditions of the evaluation.

***Examining methods*:** Finding rationale for selecting particular methods and tools serves to enhance or inhibit *multicultural validity*. Scrutinizing selected methodology will make us more familiar with a variety of techniques, thereby enhancing our ability to distinguish, detect, and control measurement bias in testing and other assessment tools.

Based on the above described methodology we will adopt a holistic evaluation approach [63] with the objective of the evaluation exercise is to assess the interfaces overall rather than in detail. Apart from the Usability problems that may have an effect on the outcome but identifying them *per se* is not the purpose of the holistic approach. Instead the emphasis is on the collection of user opinions and attitudes.

We will collect data of this kind is typically gathered by interviews or surveys and will prefer to collect most of the data using semi-structured interviews. We prefer to use semi-structured interviews over interviews or survey based on closed-ended questions because though data collection through closed-ended interviews or surveys is easy to collect and analyze, it lacks the richness of detail required when ascertaining user preferences and the reasons for them. But unstructured data is much harder to analyze.

The literature on usability evaluation recognizes the value of the qualitative data

obtained from interviews where the questions are open-ended. Unstructured data can be interpreted anecdotally because its purpose is to provide new ideas; it is well suited for exploratory studies where users can explain themselves in depth and allows user preferences to be determined. The results do not, however, lend themselves to more formal data analysis; are an indirect way of finding out about an interface and may be difficult to analyze. We have added a sample questionnaire from [64] in the appendices (Appendix D). The proposed multicultural framework [65] will then be utilized to analyze the collected data that will assist in the following-

- Categorizing the data
- Exploring data
- Theory building & concluding suggestions

3.4.4 Logistical Plan for Multicultural Evaluation

For the purpose of multi-cultural evaluation (MCE), we have identified three different categories of target groups each of which will focus on three different types of tasks as shown in the table below:

Table 11. MCE: Evaluators of targeted Focus Groups

Task No.	*Task Description*	*Evaluators*
1	Inclusion in System requirements, design and implementation through Cultural Heritage Ontology	MOSAICA SHU Team members
2	Pilot Study	Several groups of PhD students
3	Expert Study	Experts users/developers of cultural artefacts

Task 1, which concerns inclusion of multicultural notions in the system implementation through cultural heritage ontology will be conducted by domain experts of different cultures and faiths and ontology engineers. Below is a tentative list of few of them.

Table 12. Participants for Evaluating Cultural Heritage Ontology

No	Name	Profile
1	Dr. Jawed Siddiqi	Professor of Software Engineering
2	Dr. Simon Polovina	PhD in Computer Science, expert in smart applications and ontology
3	Dr. David Love	PhD in Computer Science, expert in knowledge representation and comparative study of religions
4	Dr. Richard Hill	PhD in Computer Science

Our tentative list of pilot study participants includes PhD students from different ethnic, cultural and religious/faith background. The list below contains the detail of a few of them but is not limited to them only.

Table 13. Pilot Study Participants

No.	Name	Role	Affiliation	Background Detail
1	Khalid Nasir Al-Zoubi	PhD Student	SHU	Students from multi-national Muslim Cultural Background
2	Charilaos Tsiropoulos	PhD Student	SHU	
3	Mohammed Hjoujbtoush	PhD Student	SHU	
4	Naser Aniba	PhD Student	SHU	
5	Jafar Alqatawna	PhD Student	SHU	

Experts in different fields will be invited to participate in MCE: expert study sessions. The table below shows a tentative list.

Table 14. Expert Study Participants

No.	Name	Role/Affiliation
1	Dr. Inderjit Bhogal	Methodist minister, Director of the Yorkshire and Humber Faiths Forum
2	Barbara Burford	Freelance consultant at Dept of Health
3	Bhupendra Manku	Director of Language Team and School Teacher
4	Anandi Mehmood	TBA
5	Barbara Walsh	Chief Exec. of South Yorkshire Workforce Development; NHS (Lead role for Multi-faith/Interfaith)
6	Rajesh Dhimar	Teaching and Learning, SHU
7	Dr. Mona Siddiqui	Professor of Islamic Studies and Public understanding, University of Glasgow
8	Mahmood Elsayess	President, ReadVerse.com
9	Rev. Prof. Ian Draffan	Christian Canon and Professor of Computing
10	David Love	TBA
11	Mehri Nicknam	Executive Director, Joseph Interfaith Foundation, member of Imams and Rabbis Committee and Academics & Theologians Roundtable at the department of Communities and Local Government and a regular contributor to the BBC World Service.

Participants in each of the focus groups will be invited to a presentation session to provide hands on exploration of MOSAICA with the aim to demonstrate aims, objectives and overview of MOSAICA system. In the next phase, the participants will be invited to use the MOSAICA system at their liberty and any help and support as requested by the participants will be provided. In the third stage participants will be invited to comment on their experiences based on an open-ended questionnaire, a sample of which has been provided in Appendix D and E.

4. REFERENCES

1. Quality Management and Quality Assurance Vocabulary-ISO 8402:1994. Available online from http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=20115

2. Converge Project TR 1101 (1997) "A Proposal for a Revised Transport Telematics Functions List" EU Fourth Framework Programme (FP4) – Transport Sector Project.

3. TRACAR-CEC/TTAP Project TR1059 (1996), "TRACAR Draft Validation Plan", Deliverable Number 05/40-05/45

4. IEEE Guide for software verification and validation plans (Guide 1059)

5. IEEE standard for software verification and validation plan (standard 1012)

6. Akhgar, B, Naderi M. Siddiqi J (2003) "GRACE User Requirements Specification".

7. Software Productivity Consortium http://www.software.org/PUB/VNV/hierverification.asp#The%20V%20Model

8. Akhgar, B. and Hariri, H. (2004); "Knowledge oriented strategy and Information Systems Development" An integrated Approach, SGAMC publication Vol. 4, Issue 8.

9. Akhgar, B (2001); A methodology method for Analysis and Design of complex systems. VIDA Publication company, ISBN: 964-91700-7-3.

10. GDPM Systems http://www.gdpmsystems.com/

11. Bug Tracking System http://savannah.cern.ch/projects/MATCH

12. Dori, Y.J. (2007). Educational reform at MIT: Advancing and evaluating technology-based projects on- and off-campus. Journal of Science Education and Technology, 16(4).

13. Preskill, H. (2004). The transformational power of evaluation: Passion, purpose and practice. In M. Alkin (Ed.) *Evaluation Roots*. Thousand Oaks, CA: Sage.

14. http://www.socialresearchmethods.net/kb/contents.php

15. Patton, M. Q. (1990). Qualitative Evaluation and Research Methods (2nd ed.). Newbury Park, CA: Sage Publications, Inc.

16. Strauss, A., & Corbin, J. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park, CA: Sage Publications, Inc.

17. B. Kaplan and D. Duchon, "Combining Qualitative and Quantitative Methods in Information Systems Research: A Case Study," *MIS Quarterly*, vol. 12, 1988, pp. 570-586.

18. Cross T., Bazron, B., Dennis, K., & Isaacs, M. (1989). Towards a Culturally Competent System of Care, Volume I. Washington, D.C.: Georgetown University Child Development Center, CASSP Technical Assistance Center.

19. Lee, A.S. (1994) Electronic mail as a medium for rich communication: an empirical investigation using hermeneutic interpretation. MIS Quarterly, June, 145-157.

20. Markus, M.L. (1994). Electronic mail as the medium of managerial choice. Organization Science, 5, 4, 502-727.

21. Patricia Y. Martin and Barry A. Turner. Grounded Theory and Organizational Research. The Journal of Applied Behavioral Science, Vol 22, No.2, pp 141-157 (1986).

22. Brannen, J. (2005). "Mixing Methods: The Entry of Qualitative and Quantitative Approaches into the Research Process." International Journal of Social Research Methodology (8), 173-184.

23. Barak, M., Rafaeli, S. (2004). Online question-posing and peer-assessment as means for Web-based knowledge sharing. International Journal of Human-Computer Studies, 61(1), 84-103.

24. Beichner, R., Wilkinson, J., Gastineau, L., Engelhardt, P., Gjertsen, M., Hazen, M., Ritchie, L., Risley, J., 2000. Education research using Web-based assessment systems. Available: http://www.physics.ncsu.edu/people/faculty.html (2002, July).

25. Hamilton, L.S., Klein, S.P. & Lorie, W. (2000). Using Web-Based Testing for Large-Scale Assessment. Santa Monica: Rand Education.

26. Rafaeli, S. & Tractinsky, N., 1991. Time in computerized tests: a multi-trait multi-method investigation of general knowledge and mathematical reasoning in online examinations. Computers in Human Behavior 7, 123–142.

27. Bosnjak, M. And Tuten, T.L. (2001). Classifying response ehaviours in web-based surveys, Journal of Computer Mediated Communication, 6(3).

28. Barak, M. And Dori, Y.J. (2005). Enhancing undergraduate students' chemistry understanding through project-based learning in an IT environment. Science Education, 89(1), 117-139.

29. Barak, M. (2007). Transitions from traditional to ICT-enhanced learning environments in undergraduate chemistry courses. Computers & Education, 48(1), 30-43.

30. Dori, Y.J. & Belcher, J.W. (2005). How does technology-enabled active learning affect students' understanding of scientific concepts? The Journal of the Learning Sciences. 14(2), 243-279.

31. Graesser, A.C., Person, N., & Hu, X. (2002). Improving comprehension through discourse processes. New Directions in Teaching and Learning, 89, 33-44.

32. Guzdial, M. & Turns, J. (2000). Effective discussion through a computer-mediated anchored forum. The Journal of the Learning Sciences, 9(4), 437–469.

33. Collison, G., Elbaum, B., Haavind, S., & Tinker R., (2000). Facilitating online Learning Effective Strategies for Modeertors Atwood Publishing.

34. Bogdan, R. C., & Biklen, S. K. (1982). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn and Bacon, Inc.

35. Lofland, J. & Lofland, L. H. (1984). *Analyzing social settings*. Belmont, CA: Wadsworth Publishing Company, Inc.

36. Volle, L.M. (2005). Analyzing oral skills in voice e-mail and online interviews. Language Learning & Technology, 9(3), 146-163.

37. Haigh S., Megarity J. (1998). "Measuring Web Site Usage: Log File Analysis," Network Notes #57, National Library of Canada, August 4, 1998, <http://www.nlc-bnc.ca/pubs/netnotes/notes57.htm>.

38. Anderson, S., Willard, T., Creech, H. & Bakker, D. (2001). Tools for Assessing Website Usage. Winnipeg: Institute for International Sustainable Development (IISD).

39. Turner S. (2000). "How the Web Works," May 31, 2000, Analog, http://www.analog.cx/docs/webworks.html>.

40. Ten Dam, G.T.M., & Volman, M.L.L. (2004). Critical thinking as a citizenship competence: teaching strategies. Learning and Instruction, 14 (4), 359-379.

41. Hare, W. (1979). Open-mindedness and Education. Montreal: McGill-Queen's University Press.

42. Gardner, P. (1996). Four anxieties and a reassurance: Hare and McLaughlin on being open-minded. Journal of Philosophy of Education. 30 (1), 271-276.

43. Hare, W. & McLaughlin T. (1998). Four anxieties about openmindedness: Reassuring. Peter Gardner. Journal of Philosophy of Education. 32(2), 283-

293.

44. Jonassen, D.H., Carr, C. & Yueh, H. (1998). Computers as mindtools for engaging learners in critical thinking. TechTrends, 43(2), 24-32. Available at: http://tiger.coe.missouri.edu/~jonassen/Mindtools.pdf February 2004

45. Facione, P.A., Facione, C.F., & Giancarlo, C.A., (1996). The California Critical Thinking Disposition Inventory Test Manual. Millbrae, CA: California Academic Press.

46. Dumas, J.S. & Redish, J.C. (1999). *A Practical Guide to Usability Testing*, Intellect Books, Exeter, England.

47. Jerz, D.G. (2002). *Usability Testing: What is it?* Available online from http://jerz.setonhill.edu/design/usability/intro.htm

48. Mayhewm D.J. (1999). *The Usability Engineering Lifecycle: A Practitioner's Handbook for User Interface Design*, Morgan Kaufmann, London.

49. Rubin, J. (1994). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*, Wiley, New York.

50. Nielsen, J. & Landauer, T.K. (1993). A mathematical model of the finding of usability problems, *Proceedings of ACM INTERCHI'93 Conference*, Amsterdam, The Netherlands, pp. 206-213. http://www.useit.com/alertbox/20000319.html

51. Diane L. Adams (Ed.). (1995). Health issues for women of color: A cultural diversity perspective. Thousand Oaks: SAGE Publications.

52. Texas Department of Health, National Maternal and Child Health Resource Center on Cultural Competency, 1997.

53. Patton, M. Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation*, *22*, 329–33

54. Fort, Martínez, & Mukhopadhyay, 2001 Fort, L., Martínez, B. L., & Mukhopadhyay, M. (2001). *Integrating a gender dimension into monitoring and evaluation of rural development projects.* Washington, DC: World Bank.

55. Patton, Jack T. 1994. "The philosopher's stone: a brief history of the Western tradition of education as it affects Papua New Guinea." 29(1): 3-13.

56. Mertens, 1999 Mertens, D. M. (1999). Inclusive evaluation: Implications of transformative theory for evaluation. *American Journal of Evaluation, 20*(1), 1-14.

57. Seigart & Brisolara, 2002 Seigart, D., & Brisolara, S. (2002). *Feminist evaluation: Explorations and experiences* (New Directions for Evaluation, Vol. 96). San Francisco: Jossey-Bass.

58. Guzman, 2003 Guzman, B. L. (2003). Examining the role of cultural competency in program evaluation: Visions for new millennium evaluators. In S. I. Donaldson & M. Scriven (Eds.), Evaluating social programs and problems: Visions for the new millennium pp. 167-182

59. Freire, P. (1973). Education for critical consciousness. New York, Seabury.

60. Erickson, E. (1956). The problem of ego identity. *Journal of American sychoanalytic Association 4(56),* 121.

61. Cowles, T. B. Cowles. Evaluation Methodology. 2005. The Evaluation Exchange, vol. XI no.2

62. Kemp, E.A.; Ots, S. Evaluating user interfaces using techniques from qualitative data analysis. International Conference Software Engineering: Education and Practice, 1998. Proceedings. 1998. 26-29 Jan. 1998 Page(s):222 – 229

63. Ravden, S. J. And Johnson,G. I. Evaluating usability of human-computer interfaces: a practical method, Ellis Horwood, Chichester, 1989.

64. Dey, I. *Qualitative Data Analysis*, Routledge, London, 1993.

5. APPENDICES

Appendix A – The Open-mindedness Questionnaire