

This project is funded under the
ICT Policy Support Programme part of the
Competitiveness and Innovation Framework Programme.

Project

Project Acronym:	AthenaPlus
Grant Agreement number:	325098
Project Title:	Access to cultural heritage networks for Europeana

Deliverable

Deliverable name:	Evaluation framework for the pilots
Deliverable number:	D6.1
Delivery date:	May 2014
Dissemination level:	Public
Status	Final
Authors (organisation)	Maria Sliwinska, Sebastian Michalek (ICIMSS); Marc Aguilar (I2CAT)
Contributors (organisation)	Pau Pamplona, Sergi Fernández (i2CAT), Claire Loucopoulos (Dedale)
Reviewers (organisation)	Barbara Dierickx (PACKED vzw), Corinne Szteinszneider (MCA)

Revision History

Revision	Date	Author	Organisation	Description
V0.1	2014-04-18	Maria Śliwinska, Sebastian Michalek	ICIMSS	First draft
V0.2	2014-05-08	Maria Śliwinska, Sebastian Michalek	ICIMSS	Slightly improved draft
V0.3	2014-05-14	Marc Aguilar	I2CAT	Added section 'Phases of evaluation'
V0.4	2014-05-26	Maria Sliwinska	ICIMSS	Expanded Summary, added Conclusion
V0.5	2014-05-27	Marc Aguilar, Pau Pamplona, Sergi Fernández	I2CAT	Finalised methods sections and rewrote text paragraphs to finalise document
V0.6	2014-05-28	Barbara Dierickx, Corinne Szteinszneider	PACKED/MCA	Internal review
V1.0	2014-05-30	Maria Sliwinska, Marc Aguilar	ICIMSS/I2CAT	Final document
V1.0	2014-05-30	Maria Teresa Natale	ICCU	Final check

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Project Coordinator: Istituto centrale per il catalogo unico delle biblioteche italiane
Address : Viale Castro Pretorio 105 – 00185 Roma
Phone number : +3906 06 49210 425
E-mail : info@athenaplus.eu
Project WEB site address : <http://www.athenaplus.eu>

Table of Contents

1	EXECUTIVE SUMMARY	2
2	INTRODUCTION	3
2.1	Background.....	3
2.2	Role of this Deliverable in the Project.....	3
3	EVALUATION METHODOLOGIES	4
3.1	Literature review	4
3.2	Testing methods for AthenaPlus applications	6
4	PHASES OF EVALUATION.....	7
4.1	Pilot phase 1. Internal testing	8
4.1.1	<i>Initial training and usability assessment</i>	8
4.1.2	<i>Advanced usability testing with test templates</i>	8
4.1.2.1	Test template – Architecture	9
4.1.2.2	Test template - Tools for editing content.....	10
4.1.2.3	Test template - Features of the Semantic Content Management	11
4.1.2.4	Usability questionnaire – the System Usability Scale (SUS)	11
4.1.2.5	Test template – Usability.....	12
4.1.3	<i>Software issue report</i>	13
4.1.4	<i>In-depth online interview</i>	14
4.2	Pilot phase 2. Small-scale evaluation.....	15
4.2.1	<i>On site questionnaire directed to the cultural heritage institutions</i>	15
4.2.2	<i>Interview guide</i>	19
4.2.3	<i>Short questionnaire for end users</i>	19
4.2.4	<i>Pilot event summary card</i>	21
4.3	Phase 3. Large-scale user experience evaluation	23
4.3.1	<i>Backend analysis of web analytics</i>	23
4.3.2	<i>In-app online questionnaire</i>	24
5	CONCLUSIONS	25
6	APPENDIX: SELECTED REFERENCES	26

1 EXECUTIVE SUMMARY

This deliverable coordinated by ICIMSS prepares the definition of a shared methodology to evaluate and test the creative applications for the (re)use of digital cultural heritage content, focused on: virtual exhibitions, digital storytelling, tourist routes and education referred to as the pilots. It is the first step, with the objective to develop a common method to evaluate the performance of each pilot action prepared in the framework of the WP5. The results of this deliverable will also be used in Task 6.3, which is devoted to user testing. A shared methodology will be implemented by the AthenaPlus partners, during the work devoted to evaluation and testing of the tools and solutions designed and developed by WP5. It will also help to evaluate and test existing shareware and freeware tools, already present on the web or previously developed by the project Partners in a different environment.

This deliverable comprises of the four main parts: Introduction, Evaluation Methodology, Phases of Evaluation and Conclusions. In the Introduction a background is given, and the role of this deliverable is described. The Evaluation Methodology chapter starts with a review of literature devoted to methodologies of software assessment. Among them are surveys, heuristics, interviews, laboratory usability tests with eye-tracking and think-aloud requirements, and synchronous or asynchronous remote usability tests.

From a number of methodologies used for software assessment, three methods have been chosen as the most adequate for the AthenaPlus pilots: questionnaires, interviews, and formal feedback gathered during workshops organised by the AthenaPlus partners in their countries. To familiarise users with the AthenaPlus software, a number of workshops organised are directed to a) the direct AthenaPlus partners, b) cultural heritage institutions from in- and outside the consortium, and c) to the end users, namely school teachers, university students, Third Age University members, and general internauts. For the last end user group, only the online questionnaires will be established to get their opinion about the solutions offered by the project.

Testing will be divided into the three main phases: a) the first phase of development, b) throughout or further down the development progress, and c) after the task completion. On each of these phases different users will be included into the evaluation process. In phase a) the main users will be the AthenaPlus partners and cultural heritage institutions interested in using the offered software for their content re(usage). In phase b) the same actors will take part, whilst during the last (third) phase also an evaluation from the end-users is expected, to check not only the prepared software but also the way of content presentation by means of the given software. The third and final evaluation phase is centred on the validation of the AthenaPlus software tools in a large-scale uncontrolled environment. This phase's evaluation mix is comprised of the backend analysis of a set of analytics, and a compact user evaluation questionnaire seamlessly incorporated into the tested tools.

The end-users included into the second stage of the evaluation process will mostly represent visitors of the cultural heritage institutions. When the software passes the internal tests, the large scale evaluation will start and apart from the AthenaPlus partners, cultural heritage institutions, and their clients, we will also include and engage teachers, students, Third Age University members, tourist agencies and the Internauts who represent the broadest and a very diverse type of users. In this section, procedures to report direct feedback to WP5 are also included. During the evaluation, a usability test will be put into practice, divided into two stages: the initial one with a few basic questions, and the advanced one where also systems architecture, tools for editing the content, and features of the Semantic Content Management infrastructure will be checked.

2 INTRODUCTION

This deliverable, coordinated by ICIMSS, presents the definition of a shared methodology necessary for testing and evaluation of the pilots prepared in the framework of WP5 and WP6. The pilots present the creative applications for the (re)use of digital cultural heritage content, focused on virtual exhibitions, digital storytelling, tourist routes and educational use.

2.1 Background

The goal of this task is to set up the criteria and procedures through which the available software tools can be assessed. Throughout this deliverable, the words “software tools” encompass and refer to all six use cases addressed by the software tools developed in the AthenaPlus project. These are MOVIO virtual exhibitions, MOVIO education, MOVIO touristic routes, MOVIO GIS, MOVIO mobile applications, and tool for digital storytelling. Similarly, the term “digital heritage project” refers to the specific instances of finished cultural products generated through the usage of the AthenaPlus software tools.

When working on this we have to bear in mind at least two different groups of users of the prepared software:

- On the one hand, a group of cultural heritage institution members interested in the software implementation at their work environment
- On the other hand, the end users interested in the digital exhibitions, digital storytelling, tourist routes and education already designed so their expectations can be more complex. In this case evaluation must combine the concepts of education, web navigation design, and user satisfaction.

2.2 Role of this Deliverable in the Project

The work done within this deliverable should prepare tools helping to point out strengths and weaknesses of the software designed in the framework of WP5. The results will be used within the development’s iterative process.

This deliverable will serve as a base for the WP6 Task 6.3: Running other pilots, especially the pilot devoted to virtual exhibitions and it will also help in Task 6.2: Running the pilots: Digital storytelling. Results of the aforementioned tasks will have a direct input on Task 6.4: Manual of good practices / implementing the pilots.

This deliverable is closely related to WP5 as the chosen methodology and tools prepared and used for software testing will inform the technical developers about the software weaknesses and any necessary revisions or improvements of the solutions developed by this WP.

3 EVALUATION METHODOLOGIES

3.1 Literature review

There are quite many approaches to evaluation methodology that adequately address the complexities of a software tool. According to Miller (2004), such methodology must be devised as questions that go beyond usability testing or e-learning testing and incorporate aspects of leisure and user satisfaction.

In the GLAM¹ sector, evaluation methodologies must put attention to the process of conducting visitor studies before exhibitions are developed (front-end user studies), after mock-ups of the displays have been produced (formative evaluation), after exhibition opening to the public (remedial evaluation), and after it has been closed to determine its success and to plan for future exhibitions (summative evaluation). However this methodology can't be implemented in the online environment as the nature of digital exhibitions is different because websites can be changed and updated more readily than physical exhibitions.

For the work in WP5 the most important will be to check the usability of the software. Usability definition by standards has been described as the following:

- ISO/IEC 9126-1, 2000 - "The capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions."
- ISO9241-11, 1998 - "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use."
- IEEE Std.610.12-1990 - "The ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component".

It is also underlined in the literature that the environment in which the testing takes place influences the methodology choice. Researchers discuss the effectiveness of evaluating e-learning material by bringing teachers into the laboratory for interviews and testing, versus the effectiveness of evaluating e-learning material by observing its use in the classroom. In the AthenaPlus pilots, a balanced combination of both approaches will be pursued, by blending in the evaluation mix both feedback-generating actions to be performed in controlled settings with actions carried out in real-world environments.

Surveys, heuristics, interviews, laboratory usability tests with eye-tracking and think-aloud requirements, and synchronous or asynchronous remote usability tests are all options of evaluation of the tested software. In the table below there is an overview of common evaluation methods and approaches:

<i>Evaluation method</i>	<i>Short description²</i>
AttrakDiff	Assess the user's feelings about the system with a questionnaire. In AttrakDiff questionnaire, both hedonic and pragmatic dimensions of User experience (UX) are studied with semantic differentials.
Day Reconstruction Method	A retrospective diary method in which respondents revive memories of the previous day by constructing a diary consisting of a sequence of episodes. They describe each episode by answering questions about the situation and about the feelings that they experienced.

¹ GLAM stands for galleries, libraries, archives and museums sector.

² The descriptions are taken from the Wikipedia and AllaboutUx websites.

Diaries	Participants keep a diary in which they report their experiences of product use, and aspects related to product use over a period of time
Focus groups	A focus group is a form of qualitative research in which a group of people are asked about their perceptions, opinions, beliefs, and attitudes towards a product, service, concept, advertisement, idea, or packaging. Questions are asked in an interactive group setting where participants are free to talk with other group members
Interviews	Face to face or online interview. Ask people what they think, feel, experience
Logging	Gathering data and metadata on technical events/triggers relevant to user activity
Observation	Observing what people do/experience, in the context-of-use
Questionnaires	A research instrument consisting of a series of questions and other prompts for the purpose of gathering information from respondents.
Usability@90mph	Quick usability method in which the goal is to spend ten minutes analysing the site to be evaluated and determining representative tasks, ten minutes running the tasks with representative users, and ten minutes analysing the results with the help of the audience (Marty & Twidale 2005)
User experience (UX) Curve	The participant draws one or more curves to describe how the experience about a product has changed over time. The curve drawing area is formed of a timeline and a horizontal line that divides positive and negative experiences
User experience (UX) Laddering	UX Laddering is an adapted interview method and adapted data analysis process for investigating the user experience to identify and understand the linkages between key perceptual elements across the range of attributes, consequences and values. Therefore, UX Laddering helps researchers and designers understand how concrete product attributes benefit personal values for end users.

From this range of commonly used methodologies for software assessment and testing, three methods were chosen as the most adequate for the AthenaPlus pilots: questionnaires, interviews, and observations, in the shape of feedback gathered during workshops organised by the AthenaPlus partners in their countries.

3.2 Testing methods for AthenaPlus applications

In order to provide feedback to the software tools (MOVIO virtual exhibitions, MOVIO education, MOVIO touristic routes, MOVIO GIS, MOVIO mobile applications, or Digital storytelling) providers (WP5), the first discussions lead to a broad agreement on the methods and the timings of the evaluation of the pilots to be run in the framework of WP6. Thus, in a series of online discussions and dedicated meetings during the March 2014 plenary meeting in Bucharest, it was agreed between the parties involved in WP5 and WP6 in the testing of the software tools that the following methods were to be applied for evaluation:

1. To obtain feedback from a number of workshops, which have a dual role in evaluation and dissemination
 - a. Workshops for the cultural heritage institution professionals, organised by the AthenaPlus project partners in their respective countries using the following methods
 - i. Questionnaires
 - ii. Interviews
 - iii. Informal feedback gathering (i.e. opinions collected with casual conversations, or insights generated at round tables or conferences)
 - b. Workshops organized for the end users interested in the offered by the AthenaPlus software
 - i. For school teachers
 - ii. University students
 - iii. Third Age University members

For all groups of the end users the same tools will be used, i.e. questionnaires and interviews to get their feedback, however more attention will be put here also to the content.
 - c. In addition, if necessary and accepted by partners in order to get also feedback from the experts who cannot spend time attending lengthy workshops, the structured group discussion can be issued using Delphi or focus group approaches.
 - d. An online questionnaire for the Internauts visiting the exhibitions, created on one of the free, dedicated survey sites, for example:
 - i. Free Online Survey (www.freeonlinesurvey.com)
 - ii. SurveyMonkey (<https://www.surveymonkey.com/>)
 - iii. Kwiksurveys (<http://kwiksurveys.com/>)
 - iv. Google Forms (<http://www.google.com/google-d-s/createforms.html>)

Sites will be assessed and best kind of survey technology will be chosen.

The questionnaires used during the workshops will be differentiated among the two different groups: a) professionals interested in using the AthenaPlus software tools for the creation of digital heritage projects, and b) the end users accessing the digital heritage projects as learning tools, which adequately addresses the complex interplay of usability, education, and leisure present.

In the remaining sections of this document, the possibilities identified in the literature review are further concretised and operationalised to become a fully fleshed-out evaluation framework.

4 PHASES OF EVALUATION

The AthenaPlus software tools will be tested in the three steps: a) at the first phase of development, b) throughout the further development progress, and c) after completion of the task. The software will be tested by AthenaPlus partners included in WP6, by professionals gathered at the planned national events and dedicated workshops organised by the AthenaPlus partners independently in their respective countries, and by the end users of the digital heritage productions.

At the first software presentation a simple survey of software usability will be conducted among the participants of the workshop. The main purpose will be to check if they find this software interesting and useful.

The second phase of testing will take place when the project partners, including the cultural heritage institutions cooperating with, manage to make a practical use of the offered software, so it will be tested in practice. This should bring some new desires not identified during the workshops yet. For this more advanced phase of testing the software, four questionnaires and supporting additional actions have been prepared. In this phase the system architecture, tools for editing content, semantic content management and more in deep software usability will be tested.

The third testing phase will happen when the end-users of cultural heritage institutions will be offered with the ready products, which will be the final evaluation of the software and the content, in an open and large-scale manner. The end-users will be asked to fill a simple questionnaire and add comments.

Table 1 Diagram of the three pilot phases in AthenaPlus

Pilot phase	<i>Phase 1</i>	<i>Phase 2</i>	<i>Phase 3</i>
Scope	Internal	Small-scale	Large-scale
User profiles	Pilot-running professionals	Cultural heritage professionals and end users	End users
Evaluation focus	Usability and technical evaluation	Usability, and user experience	User experience
Indicative number of users	4-5 users	20-30 users	1000+ users

4.1 Pilot phase 1. Internal testing

This phase will be divided into two steps: an initial familiarisation with the software, and in depth usability testing.

4.1.1 Initial training and usability assessment

In the initial phase of the pilots, a training will be organised for each AthenaPlus partner running a pilot. These kickoff trainings may also involve a number of external cultural heritage institutions, which are invited by the hosting institution to participate in the training in order to get acquainted with the offered tools.

During the workshops partners will familiarise with the basic concepts of a digital heritage project (for example, notions such as digital exhibition, semantic content management, and others), and more specifically with the software enabling these performances. In these trainings, the AthenaPlus software tools will be presented, and a number of examples of finished projects will be displayed to show in practice the possibilities of feature customisation. At the workshop's end, participants will be asked to fill in a simple questionnaire. This questionnaire provides the project with the first feedback.

Evaluation of the [insert name of software] tool

How do you rate:

(1 = I like it / 5 = I do not like it)

The back-end interface	1	2	3	4	5
The front-end possibilities	1	2	3	4	5

Please clarify your answers or suggest improvements to back- and front end interface:

**Have you encountered any bugs in [insert name of software] while working (briefly) with it? YES
– NO**

Please describe them below.

Based on the training session and the overview of what [insert name of software] is capable of, what would you suggest in general terms of improvement?

4.1.2 Advanced usability testing with test templates

As the next step of usability testing, the partners focus their efforts in creating their digital exhibitions/ tourist routes or educational materials with the AthenaPlus tools. In parallel, as they work with the tools, their insights and recommendations are collected and processed, so that the prepared tools can be improved continuously by the WP5 in an iterative process of problem finding – problem solving – solution implementation – solution testing. This period of internal usability testing with partners has the goal to improve the AthenaPlus software tools interfaces and functionalities from the professional user's

point of view. As an outcome, actionable user insights are generated, that make the software tools more usable and satisfying for cultural heritage professionals.

The pilot phase 1 involves four test templates and a usability questionnaire that have to be filled in by the partners using the tools in practice, a procedure to report direct feedback to WP5, and a final online in-depth interview.

Towards the end of phase 1, the partners which have been using the software in a creative way will have a good idea of the strengths and weaknesses of the tested tool. At this point, they will be asked by the WP6 lead to fill in four short templates and a standard usability questionnaire. These test templates will be complemented by the System Usability Scale, a standardised 10-item questionnaire which yields a value between 0 and 100 which encapsulates the ease of use and learnability of a given system (Brooke 1996). The purpose of the administration of this questionnaire is to get a quantitative assessment of the tool's usability, and obtain a first usability benchmark against which to compare all future improvements.

The test templates for Architecture, the Tools for editing content, the Features of the Semantic Content Management and the System Usability Scale questionnaire will be sent by the WP6 lead and answered directly by the pilot partners.

The usability template, on the other hand, requires a more complex methodological approach, because it cannot be administered directly to the user. This usability template is to be filled by a trained researcher during a task-based testing session with a small group of users. Typical numbers of users for usability tests are 4-5 representatives of the target demographic, but larger number of users (up to 10) will yield more robust results (Faulkner 2003). The users try to perform a predetermined set of typical tasks with the software tool, while they voice out their thoughts on their current experience with the tool (think-aloud protocol).

Given the technical complexity of this evaluation method, it is considered optional, and is only added here as a possibility. Pilot partners expressing interest in conducting this evaluation with a small user group will be supported in the process by WP6, and given additional methodological guidance on how to conduct usability studies.

4.1.2.1 Test template – Architecture

Tested features	Excellent	Good	Average	Needs improve ment -	Comments
Layout					
Layout - adaptation to the device (laptop)					
Layout - adaptation to the device (tablet)					
Layout - adaptation to the device (smartphone)					
Templates					

AthenaPlus D6.1. Evaluation framework for the pilots

Templates - customisation - modification possibilities					
Templates - color modification					
Templates - translation to national language					
Templates - feedback form					

4.1.2.2 Test template - Tools for editing content

Examples of the tested features	Excellent	Good	Average	Needs improvement	Comments
Map management					
Timeline					
Images					
Photogallery					
Slideshow					
Page-flip, index mechanisms, menus					
Built-in text editor and formatting tools					
Storytelling					
Documentation and help files					
Export/import tools					

4.1.2.3 Test template - Features of the Semantic Content Management

Tested features	Excellent	Good	Average	Needs improvement	Comments
Page content management - generation of pages linked semantically					
- Integration with other resources					
- Search interface					
- Browsing					
- Zooming					
- Navigation					

4.1.2.4 Usability questionnaire – the System Usability Scale (SUS)

How do you feel about the following statements regarding the application?					
	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I would like to use the application frequently					
I found the application unnecessarily complex					
I found the application easy to use					
I think I will need the help of a technician to use the application					
I found that the various features of the application were well organized or incorporated					
I found that there were too many inconsistencies in the application					
I imagine that most people will					

AthenaPlus D6.1. Evaluation framework for the pilots

learn quickly to use the application					
I found that the application was painful to use					
I felt confidence when using the application					
I had to learn many things before I could start using the application					

4.1.2.5 Test template – Usability

	Value	Comments
1. Effectiveness	Overall (1-5):	
a. Percentage of tasks accomplished during the evaluation period		
b. Percentage of failures		
2. Efficiency	Overall (1-5):	
a. Number of good characteristics recalled by users		
a. Number of bad characteristics recalled by users		
b. Time spent on one task average		
c. Time spent on errors handling and correction		
d. Documentation or help's use frequency		
e. Number of available functions and features not called upon		
3.Satisfaction	Overall (1-5):	
a. Percentage of users' favourable comments		
b. Percentage of users'		

unfavourable comments		
c. Number of times that user expresses his frustration		
c. Rating scale for users' satisfaction with available functions		
4. Learnability	Overall (1-5):	
a. Time needed to learn basic navigation		
b. Time needed to learn advanced features		
c. Time needed to learn effective use of search engine		

4.1.3 *Software issue report*

AthenaPlus partners will be asked to fulfil the questionnaires towards the end of their pilot phase 1 (never later than month 27) and send them back to the WP6 leader, responsible for the overall assessment of the pilot actions. The results of the action will be reported by the WP6 leader to WP5. However, since the evaluation of a software tool is a long-term process, it will be possible to send an email to WP5 leader with the bug report or the suggested improvement at any time it's noticed:

<i>Software issue report</i>	
Software tool	(name of software tool)
Report n ^o	(to be filled in order of reception by WP5 representative)
Date of issue reporting	(date in which the problem report is sent)
Issue detected by	(name and email address of WP6 member sending the report)
Issue forwarded to	(name and email address of WP5 member receiving the report)
Description of issue	(short description of the issue detected, including screenshots)
Suggested solution	(optionally, the report sender may include a short statement of a proposed solution to the issue)
Action taken	(short description of remedial actions taken to solve the issue or bug)
Date of issue solving	(date in which the problem is effectively solved)

Although the use of an issue tracking system such as Jira was considered for this purpose, it was finally discarded in favour of the simple bug reporting template above, because of cost-effectiveness considerations.

4.1.4 *In-depth online interview*

At the end of phase 1, a final online call will be scheduled by the WP6 lead with every partner running a pilot. The purpose of this online interview will be to debrief the participants on the experience of creating a full project with the software tools, and get the final remarks and recommendations from a professional user perspective.

This research action will probe deeper on the key topics and problematic issues that will have been detected with the analysis of the results of the test templates and the usability questionnaire. The audio recordings of these online interviews will be analysed in WP6, and a Phase 1 wrap-up report will be produced (to be included in D6.6).

4.2 Pilot phase 2. Small-scale evaluation

In phase 2, the pilot's attention shifts to usability, and user experience evaluation with small groups of external users, which include cultural heritage institutions interested in creating their own digital projects with the AthenaPlus software tools, and visitors or users of these projects.

The evaluation efforts of this phase revolve around a series of real-world actions with professionals and with end users. The goals of these actions are to dynamise the stakeholder groups, involve other institutions in working with the AthenaPlus software tools, disseminate the software tools, and collect useful feedback on usability, and user experience.

These actions are directed to a) the AthenaPlus partners, b) cultural heritage institutions, experts and professionals, and c) to the end users, namely school teachers, university students, Third Age University members, and Internauts. In this phase, it is particularly important to engage professional stakeholders and experts from the different domain areas covered by the pilots (digital exhibitions, digital storytelling, tourism, GIS and education), and elicit their qualified opinion on the software tools' usefulness, and ease of use.

Thus, whereas phase 1 was a strictly internal matter which only involved AthenaPlus partner institutions, phase 2 is all about engaging small groups of cultural heritage stakeholders interested in using the software. This is done in the framework of events organised by the pilot-running institutions, a controlled environment where qualitative feedback can be easily collected by the trainer. As the pilot proceeds and the digital heritage project is improved based on the collected feedback, more and more stakeholders may be brought into the pilot's community of users.

Evaluation of the AthenaPlus software tools organised in this phase involves a questionnaire with items that probe about several dimensions of the user experience, the usability and the professional stakeholder's perception of the software tool, an interview guide or list of relevant research questions for on-site interviews and focus groups in real-world activities.

In addition to User experience UX data, other variables such as geographic information, cultural consumption and socio-demographic variables may be included in the questionnaire and interview guide, if these are unknown to the institution organising the physical event. If needed, the questionnaires may be translated into the languages of choice of the event attendees.

4.2.1 On site questionnaire directed to the cultural heritage institutions

AthenaPlus feedback questionnaire

1. Respondent profile

Age Sex Occupation

Last week I used the Internet... Less than 4 hours 4 to 8 hours
8 to 12 hours More than 12 hours

2. Are you one of the partners in yes no- please state your
the AthenaPlus project? institution:

3. How did you hear about this event?
- | | |
|-------------------------------------------|--------------------------------------------------|
| <input type="checkbox"/> leaflets/posters | <input type="checkbox"/> social media |
| <input type="checkbox"/> newspaper | <input type="checkbox"/> email |
| <input type="checkbox"/> invitation | <input type="checkbox"/> others [please specify] |

4. How would you rate [name of tool] in terms of the following:

	Not important	Important	Very important
Relevance to your work			
Relevance to your country context			
Relevance to the tourism sector			
Relevance to the Cultural Heritage sector			

5. Which features do you find most attractive in [name of tool]?

--

6. How satisfied do you feel about [name of tool] in regards to the following aspects?

	Very dissatisfied	Dissatisfied	Satisfied	Very satisfied
Interest				
Ease of use				
Aesthetics				
Practical application				
Innovativeness				

7. How do you feel about the following statements regarding [name of tool]?

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

I would like to use it often					
I found it unnecessarily complex					
I found it easy to use					
I think that I would need technical support to use it					
I found that the various features of it are well organised and integrated					
I found that there were too many inconsistencies in					
I imagine that most people will quickly learn to use it					
I found that it was painful to use					
I felt confident when using it					
I needed to learn a lot of things before I could start with it					

8. Please share with us your opinion on the following statements:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
This product strengthens the exploitation of the cultural resources of my institution					
This product helps the promotion of a place and its territory					
The quality and attributes of this product appeals to tourist and day trippers					
The quality and attributes of this product appeals to the local population					
This product may benefit and impact on the economic					

development of the territory					
Digital technologies improve the user experience and the understanding of Cultural Heritage					
The use of digital technologies improves the distribution and consumption of Cultural Heritage					
I gained new knowledge on the uses of digital technologies applied to Cultural Heritage from this event					
I can apply this knowledge in my work and/or network					

9. Which additional functionalities do you think should be added to [name of tool]?

9. As a whole, how satisfied are you with [name of tool]?

Not at all 1 2 3 4 5 6 7 8 9 10
 extremely likely

10. How likely is it that you would recommend [name of tool] to others?

Not at all 1 2 3 4 5 6 7 8 9 10
 extremely likely

11. Do you have any other comments or suggestions that you would like to share with us?

Thanks for your valuable feedback!
 The AthenaPlus team

4.2.2 Interview guide

Another useful method for this pilot phase is an interview. User interviews and focus groups are more flexible than the prearranged sets of questions that are characteristic of surveys. Semi-structured or unstructured interviews give participants the freedom to elaborate on the points that are more relevant to them, and to dismiss those that are not important to their experience. Similarly, they give the researcher a chance to pursue emerging leads that could be of interest, and to probe deeper on difficult topics.

The formality of interviews can range from very formal and structured to informal and loose. Interviews can be arranged beforehand and conducted in a research setting, taped and then transcribed. Conversely, they can also be spontaneous events, capitalising on an unforeseen opportunity to obtain data from an informant. In both cases, it is important that the researcher conducts them according to a protocol (a formal document stating the research questions to be asked and providing guidance on conducting the interview), and that output is reported according to a template.

A popular variant of an interview is a group interview, or focus group. Focus groups are "carefully planned series of discussions designed to obtain perceptions on a defined area of interest in a permissive, nonthreatening environment" (Krueger & Casey, 2009). These allow participants to engage in a discussion focused on particular topics that are of interest to the researcher. Focus groups are organised to facilitate a situation where insights about users' everyday situation can be gained. Focus groups are helpful when wanting to facilitate interactive communication among newly formed conversation groups that share characteristics of interest (Ståhlbröst, 2008).

Below there is a list of some questions that can be used to direct the feedback gathering effort in interviews and focus groups with stakeholders, at workshops, trainings and other physical events:

- *What do/did you expect from this event/service?*
- *In general, how satisfied are you with this service/event?*
- *Has this event fulfilled your expectations?*
- *Is there anything you would like to change about this event/service?*
- *How easy did you find using this service?*
- *How much training do you think you would need to learn how to use this service?*
- *Is there any change that you think would make it easier to use this service?*
- *How well do you think the elements that make up this service fit in?*
- *How much does this service fulfil your needs?*
- *How innovative have you found this service?*
- *Is this the first time you use a service of this kind?*
- *How often do you think you would use this service?*
- *Would you recommend this service to your professional contacts?*

4.2.3 Short questionnaire for end users

In addition to these actions directed at professionals, feedback will also be collected from the end users attending on-site events. The structure of these events will be tailored to the profile of the selected end user group.

AthenaPlus feedback questionnaire

1. Respondent profile

Age Sex Occupation

Last week I used the Internet... Less than 4 hours 4 to 8 hours
8 to 12 hours More than 12 hours

2. How did you hear about this event? leaflets/posters social media
 newspaper email
 invitation others [please specify]

3. How satisfied do you feel about [name of digital heritage project] in regards to the following aspects?

	Very dissatisfied	Dissatisfied	Satisfied	Very satisfied
Interest				
Ease of use				
Aesthetics				
Practical application				
Innovativeness				

4. As a whole, how satisfied are you with [name of digital heritage project]?

1 2 3 4 5 6 7 8 9 10
Not happy at all extremely happy

5. How likely is it that you would recommend [name of digital heritage project] to others?

1 2 3 4 5 6 7 8 9 10
Not at all likely extremely likely

6. Do you have any other comments or suggestions that you would like to share with us?

Thanks for your valuable feedback!
The AthenaPlus team

4.2.4 *Pilot event summary card*

To facilitate focusing attention on the most relevant items and ease the reporting of the insights gained at real-world events during Phase 2 and beyond, a pilot event summary card has been created to be filled by the AthenaPlus partners:

AthenaPlus - Pilot event summary card	
Software tools tested	(describe the specific technologies used: MOVIO virtual exhibitions, MOVIO education, MOVIO touristic routes, MOVIO GIS, MOVIO mobile applications, or Digital storytelling)
Pilot country	(name of country in which the pilot is taking place)
Pilot institution	(name of institution responsible for running the pilot)
Event description	(describe the specific cultural activity taking place: museum visit, storytelling event, music concert, tourist visit to point of interest,...)
Context	(specify the context of the observation: cultural heritage conference, dissemination event, commemoration day, physical exhibition opening day, etc...)
Date and hour	(date and hour of start and closure of the event)
Place and location	(describe place and location where the event took place)
Participants	(estimated total number of participants or attendees)
Participant profiles	(describe profiles of participants or attendees: sex, age group, occupation, organisation,...)
Criteria for activity selection	(specify the reasons why this particular activity was selected to be observed)
Short description of the event	(description of the activities that took place during the event: if available, add the event agenda)
Observation topics	(list of topics that focused the attention of the researcher(s) during the observation)
Summary of event outcomes	(bullet point list of most relevant outcomes of the event)

AthenaPlus - Pilot event summary card	
Suggested recommendations for improvement of tested tool	(bullet list of suggested recommendations for improvement of the tested tool for WP5 technical team)
Reported new ideas on creative uses of tools	(bullet list of suggested use cases, strategies for user involvement or experimentation scenarios for WP6 user community team)
Comments	(room for miscellaneous comments)
Generated documents	(list of generated documents: video recordings of the event, interview or round table transcripts, photo album of the event...)

4.3 Phase 3. Large-scale user experience evaluation

In the third and final phase, an improved version of the digital heritage project is published and is open to visitors and users from all around the world.

At this point qualitative feedback becomes more problematic, as stakeholders are largely unknown and, with some of the AthenaPlus tools (i.e. MOVIO digital exhibitions), they may even be spread out in several continents. Thus, the most efficient way to collect feedback is with quantitative tools, integrated directly in the software tools.

Validation of the software tools will be achieved with a combination of backend analysis (log analysis), and a short and compact User experience questionnaire. This questionnaire pops up to each unique user after the first time he or she completes a virtual visit of a certain length (i.e. more than 5 minutes).

4.3.1 Backend analysis of web analytics

The analysis of data logs can be useful to uncover usage patterns and detect bugs to be fixed. Log analysis is a technique whose goal is to track user activities when the user is engaged in using the application, service or platform (Jansen, Spink, & Taksa, 2009, pp. 124-142). With this approach, all access and error information can be logged as user-driven events (e.g. 'User A clicks button X') and analysed later to mine useful information. The log data can be analysed using statistical tools such as descriptive statistics and when needed, extended with more qualitative tools like pattern matching or netnography (Jansen, Spink, & Taksa, 2009, pp. 488-505).

The specific kind of information to be logged may vary from software tool to software tool, depending on the particular research questions that are considered to have been insufficiently covered in previous phases. These may include the following analytics (Web Analytics Association, 2007):

<i>Metric</i>	<i>Definition</i>
Total page views	The total number of times a page (an analyst-definable unit of content) was viewed during a reporting period.
Number of unique visitors	The number of inferred individual people (filtered for spiders and robots), within a designated reporting timeframe, with activity consisting of one or more visits to a site. Each individual is counted only once in the unique visitor measure for the reporting period.
Average number of page views per session	<p>A visit is an interaction, by an individual, with a website consisting of one or more requests for an analyst-definable unit of content (i.e. "page view"). If an individual has not taken another action (typically additional page views) on the site within a specified time period, the visit session will terminate.</p> <p>Thus, page views per visit is the number of page views in a reporting period divided by number of visits in the same reporting period. Similarly, average session length is the total time spent by all users visiting the web, divided by the total number of visits.</p>
Average session length	

Percentage of repeating visitors	The percentage of unique visitors with activity consisting of two or more visits to a site during a reporting period.
Exit rate of a given page	Exit page is the last page on a site accessed during a visit, signifying the end of a visit/session. Therefore, the page exit ratio is the number of exits from a page divided by total number of page views of that page.
Average active or engagement time	Average amount of time that visitors spend actually interacting with content on a web page, based on mouse moves, clicks, hovers and scrolls.

Besides error fixing, the objective, non-contextual information obtained through the analysis of log data can be used to provide a better understanding of the actual use of services by detecting interesting cases of user behaviour, and be used for the optimization of the platform. Quantitative logging analysis can also assist in sampling users for qualitative methods by pinpointing the most information-rich cases. In that sense, a central purpose of logging analysis is to select those participants that can be monitored and examined more closely, by way of the explicit feedback capturing mechanisms such as observation, questionnaires and qualitative interviewing that have been used in phase 2.

4.3.2 *In-app online questionnaire*

The final validation of the AthenaPlus software tools will be accomplished with a short user satisfaction questionnaire. This questionnaire will be integrated into the AthenaPlus tools, in a way so that the questionnaire is displayed to every unique user at the end of the first session of a length over five minutes. Users will have the possibility to skip the questionnaire, to avoid random response patterns due to uninterested users being forced to submit a response. Repeating users may be re-administered the questionnaire every Nth time they re-access the digital heritage project, so that changes in the user's perception of the quality and usefulness of the AthenaPlus tools can be tracked over time (learning effects).

The questionnaire is limited to just two questions (10-point satisfaction scale and open-ended comment box) to reduce the user's effort in answering it to the least possible, and thus maximise the response rate.

AthenaPlus feedback questionnaire

How happy are you with your experience with [name of digital heritage project]?

1 2 3 4 5 6 7 8 9 10
 Not happy at all extremely happy

Do you have any comments or suggestions that you would like to share with us?

Thanks for your valuable feedback!

5 CONCLUSIONS

Presented in this Deliverable is the methodology of the AthenaPlus pilots assessment, which will be used by the project partners in assessment of the creative applications for the (re)use of digital cultural heritage content prepared by WP5, focused on: digital exhibitions, digital storytelling, tourist routes and education. Evaluation of the pilots will start together with the trainings focused on digital exhibitions software, whilst testing of the pilots devoted to education and tourist routes will start in month 20. The pilot actions will run based on the developed services to result in a report in M27.

Results of the assessment, based on the prepared Methodology, will be reported to the WP5 leaders to allow the final re-iterations for improvement to the developed infrastructure.

6 APPENDIX: SELECTED REFERENCES

Abdinnour-Helm, S. F., Chaparro, B. S., & Farmer, S. M. (2005). *Using the end-user computing satisfaction (EUCS) instrument to measure satisfaction with a web site.* „Decision Science”, vol.36(2), pp. 341-364.

Brooke, J. (1996). "SUS: a "quick and dirty" usability scale". In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland. *Usability Evaluation in Industry*. London: Taylor and Francis.

Griffiths, J.R. (2003) "Evaluation of the JISC Information Environment: student perceptions of services" *Information Research*, **8**(4), paper no. 160 [Available at: <http://informationr.net/ir/8-4/paper160.html>] *guidelines for creating effective surveys*. New York, NY: Customer Service Group.

Faulkner, L. (2003). *Beyond the five-user assumption: Benefits of increased sample sizes in usability testing*. Behavior Research Methods, 35(3), 379-383.

ISO FDIS 9241-210:2009. Ergonomics of human system interaction - Part 210: Human-centered design for interactive systems (formerly known as 13407). International Organization for Standardization (ISO)

Jansen, B., Spink, A., & Taksa, I. (2008). *Handbook of research on Web log analysis*. Hershey, PA: IGI Global.

Krueger, R.A., & Casey, M.A. (2009). *Focus groups: A practical guide for applied research* (4th Ed.). Thousand Oaks, CA: Sage Publications.

Marty Paul F. , Twidale Michael B. (2005) *Usability@90mph: Presenting and evaluating a new, high-speed method for demonstrating user testing in front of an audience.* "First Monday", Volume 10, No 7 - 4 <http://firstmonday.org/ojs/index.php/fm/article/view/1260/1180>

Miller, L. (2004). *User satisfaction surveys.* "Australasian Public Libraries and Information Services", vol. 17(3), pp. 125-133.

Nilsen, K. (2004). *The Library Visit Study: user experiences at the virtual reference desk.* *Information Research*, **9**(2) paper 171 [Available at <http://InformationR.net/ir/9-2/paper171.html>]

Spunt, T. M. (1999). *Guide to customer surveys: Sample questionnaires and detailed guidelines for creating effective surveys*. Purdue University Press

Ståhlbröst, A. (2008) *Forming future IT: the living lab way of user involvement*. Luleå University of technology, PhD thesis.

Web Analytics Association (2007). *Web Analytics Definitions – version 4.0.* [Available at: http://www.digitalanalyticsassociation.org/Files/PDF_standards/WebAnalyticsDefinitionsVol1.pdf]