



SALERO

Evaluation Methodology

Deliverable 9.2.1



Evaluation Methodology

SALERO Deliverable 9.2.1

SALERO identifier: SALERO-D921-UPF-EvaluationMethodology_v05.doc

Deliverable number: D9.2.1

Author(s) and company: Josep Blat, Sergio Sayago, Daniel Soto,
Juan Abadía (FBM-UPF)

Work package/task: WP09

Document status: Final

Confidentiality: Restricted

DOCUMENT HISTORY

Version	Date	Reason of change
1	2006-07-21	Initial version (v.0)
2	2006-07-25	Executive summary and extra references added, improvements in the document
3	2006-07-28	Version for internal review, with BLITZ suggestion added.
4	2006-08-01	Minor changes (typos, formatting) suggested by internal review (JRS)
5	2006-08-11	UPF final version

The work presented in this document was partially supported by the European Community under the Information Society Technologies (IST) priority of the 6th framework programme for R&D.

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content.

This document contains material, which is the copyright of certain SALERO consortium parties, and may not be reproduced or copied without permission. All SALERO consortium parties have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the SALERO consortium as a whole, nor a certain party of the SALERO consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and does not accept any liability for loss or damage suffered by any person using this information.

Table of Contents

1	Executive Summary	1
2	Introduction	2
2.1	Purpose of this Document	2
2.2	Scope of this Document	2
2.3	Related Documents	3
3	Project Objectives and Experimental Productions	4
3.1	The Objectives of the Project	4
3.2	Experimental Productions and Challenges	4
3.2.1	<i>Information and Entertainment Programming Experimental Production</i>	5
3.2.2	<i>Cinematic Content Adaptation Experimental Production</i>	5
3.2.3	<i>Pre-School Experimental Production</i>	5
3.2.4	<i>Interactive Game Experimental Production</i>	5
4	Evaluation Methods and their Role	7
5	Additional Categorisation of Methods	11
6	Categories of Evaluation Actors, Actions and Areas	13
6.1	Actors	13
6.2	Evaluation Committees	13
6.3	Actions	13
6.3.1	<i>Case studies</i>	13
6.3.2	<i>Cost Benefit Analysis</i>	14
6.3.3	<i>Focus Groups/Workshops</i>	14
6.3.4	<i>Interviews</i>	14
6.3.5	<i>Questionnaires</i>	14
6.3.6	<i>Reviews</i>	14
6.4	Areas and Dimensions	14
6.4.1	<i>New methodologies</i>	14
6.4.2	<i>Workflow improvements</i>	14
6.4.3	<i>Tools</i>	14
6.4.4	<i>Quality of content produced</i>	15
6.4.5	<i>End user satisfaction</i>	15
6.4.6	<i>Transferability</i>	15
7	Schedule of Evaluation	16
8	Conclusions	18
9	References	19

10 Glossary.....20

1 Executive Summary

This deliverable sets forth the evaluation objectives, procedures and schedule of evaluation instruments and outputs for the initial 18 months of the SALERO project, and introducing the evaluation perspectives for the whole project. This deliverable focuses on the definition of an evaluation methodology for assessing Experimental Productions in an integrated way. Namely, this deliverable deals with the following aspects:

- the main areas to be evaluated in the Experimental Productions
- the different methodologies for evaluation
- the actors in the evaluation process and the definition of an Evaluation Working Committee
- the timetable for the implementation of evaluation instruments and the production of evaluation reports

SALERO Description of Work includes experimental productions to be developed in four different categories: information and entertainment programming, cinematic content adaptation, pre-school, and interactive games. It also defines the main envisaged research challenges which need to be evaluated in each experimental production.

Due to the wide range of evaluation methods and the heterogeneity of aspects to be assessed, this deliverable selects a subset of methods, which range from inspection to user-based evaluation methods. These methods are expected to allow us to cover all aspects which need to be evaluated and validate results against evaluation criteria or metrics, which have already been defined in the project.

The level of involvement of participants in the experimental productions can vary, due to the spiral cycle development. Nevertheless, their involvement in evaluation activities is very relevant in order to achieve SALERO's goals. At this stage of the project, we define broadly the range of actors which are expected to play an active role in the evaluations of experiment productions, such as professionals in pre-production, digital content creators and managers. In addition to these actors, this deliverable suggests defining an Evaluation Working Committee (EWC), which will consist of one evaluation representative for each partner and the evaluation coordinator who will oversee the whole evaluation.

A schedule for the evaluation instruments and outputs for the evaluation of the first experimental productions is defined in this deliverable. The results of these evaluations will pave the way for the more refined and consolidated methodology to be implemented in the next development cycle.

2 Introduction

2.1 Purpose of this Document

There is a wide range of aspects in SALERO which can (and should) be assessed. However, according to SALERO DoW, our goal within this WP is to deal with the evaluation procedures for the experimental productions, providing a plan “which will include the timeline for developing the instruments (specific questionnaires, experiment plans et cetera) required for each of the assessments”. The plan will be reviewed “in the course of evolution of the project, as may be necessary”. Thus, our specific WP goal is to plan and implement an integrated evaluation of the overall objectives of the project through an assessment which “will be heavily user-centred with an emphasis on the performance of the new technologies under field trial conditions in the experimental productions”.

2.2 Scope of this Document

During the first 18 months of the project and in a natural spiral development strategy, the focus is on Prototypes of the Experimental Productions, with the following timetable

- Scenarios are expected by the end of month 8
- Specifications and Raw Media, by the end of month 12
- Prototypes and their evaluation, by the end of month 18.

The Evaluation Implementation will follow closely this spiral cycle. This document will address the methodology for the whole project, but providing more detail for the prototyping cycle until the end of month 18. There will be an Evaluation Plan document on month 13, on the basis of the Specifications of month 12, followed by an Evaluation Report on month 18, which will include updating of this Evaluation Methodology document according to the needs of the project and the results of the first period of evaluations.

In this methodology deliverable, we will address the main aspects to be evaluated in the experimental productions; the different methodologies for the evaluation; the actors in the evaluation process; the Evaluation Working Committee, and the time table for producing the specific instruments and evaluation reports, together with the updates of the evaluation plan.

According to SALERO objectives, the main metrics should be

- (a) audience satisfaction and judgment of quality of programme material;
- (b) imperceptible integration of re-used or modified material and newly-originated material (colour, resolution, movement, detail, sound quality and ambience, behaviour);
- (c) benchmark measurement of system performance and
- (d) technology usability evaluation with professional operators.

The remainder of the deliverable is organised in the following way. Section 2 aims to contextualise the evaluation by summarising the objectives of the project and the main challenges associated with the several experimental productions. Section 3 and 4 give a full description of the evaluation methods and their role in the experimental productions. Section 5 and 6 define the main actors and areas to be evaluated, along with a schedule of evaluation instruments and outputs, respectively.

2.3 Related Documents

Before reading this document it is recommended to be familiar with the SALERO Description of Work in the version of 18 July 2006.

3 Project Objectives and Experimental Productions

3.1 The Objectives of the Project

Let us remind that SALERO “aims to advance the current state of the art in digital media to the point where it becomes possible to create audiovisual for cross-platform delivery using intelligent content tools, with greater quality at lower cost, to provide audiences with more engaging entertainment and information at home or on the move” and this is translated into three global objectives:

O1: Research into methodologies for describing, creating and finding intelligent content

O2: Toolsets to create, manage, edit, retrieve and deliver content objects, addressing characters, objects, sounds, language sets, and behaviours

O3: Verifying the concept of intelligent content by experimental productions and promoting the take-up of intelligent content technologies

As we see, one of the goals of the third objective is the validation of the objectives 1 and 2 through the experimental productions. Thus, one of the key elements for the background of the evaluation is the plan for the experimental productions.

3.2 Experimental Productions and Challenges

SALERO DoW includes experimental production created in four different categories:

- Information and entertainment programming
- Cinematic Content Adaptation
- Pre-school and
- Interactive games

The experimental productions should provide the background against which “assess how media objects, audio, characters, character environments can self-adapt between various media types, genres and styles”. The DoW, identifies the main research challenges posed by the individual productions. This is the basis for our evaluation plan, although all research innovations will be applied to each experimental production at a different level.

Let us remark that the challenges identified for all of the experimental productions are:

- To create body movements database and language
- To improve skeletal animation engine with automatic import of movements
- To create a high quality flexible facial models
- To develop a multiplatform OpenGL-based graphics engine, with high-quality real-time rendering
- To create a search and retrieval engine for Music and Sound Effects.
- To develop software toolkits, software systems, plug-ins and interfaces that allow the control of appearances, sounds, semantic behaviour and properties of intelligent content objects for media production and post-production

In the course of defining the experimental productions, the project has to make sure that they cover the most significant issues which SALERO intends to address, so that assessment of the results can be performed, and success measures in the key aspects of SALERO can be provided.

Let us recover now the main research challenges identified in the DoW for the different experimental productions, which will provide the basis for the questions to be asked in the evaluation of the experimental productions are described in the following sections:

3.2.1 Information and Entertainment Programming Experimental Production

- To allow characters to be adapted easily to various formats, graphic presentations, and settings, developing procedures for manipulating the appearance, sound, movement and behaviour of semantically aware characters and other objects for delivery on different platforms in a range of context.
- To allow considerable automation of content.
- To provide various voice generation possibilities, meeting all production needs for static or dynamic presentations, developing improved methods and tools for script translation and speech synthesis as a means of supporting the generation of multilingual media content.

3.2.2 Cinematic Content Adaptation Experimental Production

- Image reproduction and automatic colour modification to suit the requirements of particular output media and playback devices (such as projectors, TV, laptops, portable media viewers, games consoles).
- Metadata extraction and use to modify the media characteristics and appearance appropriately for the platform, including the ability to match and the lighting characteristics of inserted characters with original scenes (and adjust them interactively).
- Translation & simplification of non-interactive, non-real-time descriptions of surface & texture into forms that can be used in interactive media with real-time rendering.
- Semantically-based, automated workflow tracking of metadata descriptions of the new output context & essence and description of the modification of the workflow needed to reflect the new purpose/use of the media.

3.2.3 Pre-School Experimental Production

- To define the search parameters for multimedia objects, so that they can appropriately be reused.
- To define the desirable properties and behaviours that objects should have when they are reused in different circumstances or environments.
- To develop an appropriate ontology language and a fully-fledged ontology management environment.
- To find suitable representation techniques for an efficient and expressive representation of semantic aspects of media components.
- To unify the worlds of metadata and expressive ontology languages.

3.2.4 Interactive Game Experimental Production

- To create techniques of procedurally generation that can be applied to the game worlds.
- To generate more realistic environments than by traditional techniques.

- To develop software toolkits, software systems, plug-ins and interfaces that allow the control of appearances, sounds, semantic behaviour and properties of intelligent content objects for media production and post-production.

4 Evaluation Methods and their Role

The number of evaluation methods is very large¹; without the aim of being comprehensive in addressing each of the possible methods we present a table of a variety of methods, with a description of each and its possible role in SALERO evaluation procedures. These methods will allow us to carry out evaluation actions according to the three metrics defined in the project (which have previously been outlined in the introduction): audience satisfaction; integration; system performance and usability evaluation.

Method	Description	Use in SALERO
Case study	A longitudinal study in a real-life context, especially when the boundaries between phenomenon and context are not clearly evident. Extensive material is gathered and a narrative developed. More effective if validated by participants, and cross-compared. In many ways Case Studies produce similar results to Ethnographic Studies, but with a wider range of inputs and less rigorous observation processes.	Case studies will be developed as reports on all the experimental productions carried out.
Cost / benefit analysis	<p>The costs of creating, implementing or using the resource to be evaluated can be assessed, and this information is useful in the construction of a business case for its adoption or continued development.</p> <p>This can consider a range of factors such how long it takes to input content, administrative workload, technical support required, hardware costs, how long it takes to program etc. It may also be desirable to establish the comparative costs with the status quo, and offset these against the benefits of the new resource. However many benefits can be intangible and hard to quantify in cost terms.</p>	Cost / benefit analysis is key for SALERO, and experimental productions are envisaged with that in mind. Cost/benefit has to be part of the evaluation reports.

¹ A comprehensive review of a wide range of methods is available at: Jacko, J.A. and Sears, A. (eds.). *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. Lawrence Erlbaum Associates, London, 2003; a classic solid reference in usability testing being: Jeffrey Rubin: *Handbook of Usability Testing. How to plan, design, and conduct effective tests*, John Wiley & Sons, New York, 1994.

Method	Description	Use in SALERO
Ethnographic study / Contextual design	<p>'In situ' observation, with the events being observed should be as little disturbed and as authentic as possible. Sources of information may include:</p> <ul style="list-style-type: none"> • Naturally occurring documentation. • Participant observation. • Audio and video recording. • Field notes and transcribed conversation. <p>A “cheap” engineering alternative of ethnographic study is Contextual Design², which is a methodology for designing, implementing and installing new solutions.</p>	<p>The strength of ethnographic study for the purposes of SALERO is that it provides an in depth rich picture of the multi-layered impact of the SALERO experimental productions.</p> <p>Ethnographic studies present two difficulties for SALERO</p> <p>a) They are very time consuming, and the results are contingent on the skill and expertise of the ethnographer.</p> <p>b) They present problems of privacy when working with professional groups.</p> <p>A key idea of Contextual Design which is the observation in the work environment is implemented in the experimental productions, and will be included as part of the evaluation.</p>
Experiment	<p>Specific design features may be investigated with a focused experiment investigating the participants' performance of specified tasks, or their perception of particular features. As a general guide experiments should use at least two groups of people with 15 participants per group if the results are to be significant.</p>	<p>At some partial aspects of SALERO this approach might be useful. It is not appropriate for complex situations such as the experimental productions envisaged.</p>
Focus groups [3]	<p>A moderated meeting of 'involved' people discussing their experience. Unlike an individual interview, a focus group enables the participants to influence each other's views as they emerge during the discussion. The advantage of this is that through disagreement and discussion, a wider range of views may emerge. The potential disadvantage is that dominant characters in the group may suppress the views of the less dominant. It is therefore normal to form groups from people of similar status.</p>	<p>Focus groups could be used effectively as a debriefing activity at the end of experimental productions.</p>
Interviews [1]	<p>Interviews are appropriate for gathering in-depth data from small numbers of people. More structured interviews make it easier to record parallel data from all interviewees, while unstructured interviews can catch responses that were not anticipated by the evaluators. They are also useful for collecting issues, which can then be measured in larger numbers with a questionnaire.</p>	<p>Semi-structured interviews will be one of the main ways in which we obtain information from experimental production participants.</p>

² Hugh Beyer, Karen Holtzblatt: *Contextual Design. Defining Customer-Centered Systems*, Morgan Kaufmann, San Francisco, 1998

Method	Description	Use in SALERO
Observations	Observations involve observation of users at work with the resource. This can be done with a checklist to record how the user gets on with particular features of the resource. Observations can be useful for highlighting specific navigation issues, or for finding anomalies in the content. One option is for the evaluator to offer him or herself as a helper so that the user will ask when there is a problem, which in itself helps to highlight issues. If the sessions are recorded on video then more detailed analysis can be conducted, and events, which were not noticed at the time of recording, may be seen to be significant.	In the complex situation of experimental productions, an extensive observation beyond the remark earlier on the role of "Contextual Design" does not seem an appropriate evaluation method, at least in the initial prototypes phase. In depth observation of some components might become relevant at later stages of SALERO.
Questionnaires [7]	<p>Questionnaires can be used to canvass very large numbers of users or potential users, and data can be gathered from each respondent in far less time than with an interview or focus group. They are generally better for unambiguous, factual data than for opinions or descriptions that require explanation or justification.</p> <p>Closed or structured questions that allow only a limited range of answers require the designers to anticipate the likely responses. This can be made easier by using the range of answers given in interviews and focus groups as a basis for questionnaire design.</p> <p>More "open" questions can be included, but these require more thinking and more time on the part of the respondent and are therefore less likely to be answered.</p> <p>Low response rates are a common problem with questionnaires and frequently results in data being gathered from a small number of self-selecting and therefore unrepresentative respondents.</p>	<p>Structured questionnaires will be used extensively. Some more open questions will also be included when administered in person to those who have taken part in experimental productions.</p> <p>While questionnaires have their limitations, the ease with which they can be administered and processed means that they provide data which would otherwise be too time consuming to gather.</p> <p>They will be dominantly used when dealing with final users (media consumers).</p>
System log data	This data is useful in tracking users' preferences and navigational choices, and can be triangulated with observations, questionnaires and interviews.	This might be an interesting option which developers of SALERO tools might consider to include for advanced versions of the tools, with orientation to very specific features of the tools. Its use would be in evaluation of final stages of the tools development for restricted and very relevant functionalities.
Textual data	Text can be harvested from email and conferencing systems and analysed. This technique might complement observation, case study analysis and other methods.	This method might be too time consuming, without providing enough added value.
User diaries	A diary kept over a period of using the product is in many ways a longitudinal think-aloud. This enables the evaluator to build up a rich picture of how the user's responses evolve over time. The diary is often a valuable complement to more structured records of use, and may provide starting points for interview discussions.	Diaries might be useful during experimental production implementation to gather more data in this complex process.

Method	Description	Use in SALERO
Reviews	External experts or peers may review the product. The quality of the evaluation depends very much on the expertise of the reviewer. It is recommended that reviewers are kept for the life cycle of the product.	Internal peer review is a method part of the SALERO project, performed by partners not involved in the specific work developed. While more independent peer review is advisable, external peers are hard to get involved in a project such as SALERO.

Again, without being exhaustive, we give some examples of significant evaluation methods which will not be used with the SALERO experimental productions, as these methods are more appropriate for the evaluation of components rather than relatively complex productions.

- **Think-alouds [2]:** The user is asked to describe his or her thought processes as he or she is using the resource; while this method is rated by Nielsen as one of the best methods in terms of results versus efforts, its use in a complex experimental evaluation environment is unfeasible. Its use for SALERO components can be envisaged.
- **Usability inspection [6]:** A set of methods based on having evaluators inspect or examine usability-related aspects of a user interface. In the context of SALERO, this method could be used for individual components.
- **Competitive analysis [5]:** This method compares the advantages and disadvantages of products/resources available. It is usually performed before the design stage begins, in order to consider features that might be interesting for the product to be developed. Sometimes it is used for summative evaluation.
- **Heuristic evaluation [4]:** this method consists of listing a set of desirable usability features, considered important in evaluation guidelines, and having an expert discuss how well the product rates against them. Again, this method is more appropriate for components rather than complex processes such as the experimental productions.

There are aspects of technical evaluation of software components which are not addressed by this document, as the workpackage is concerned with evaluation at the project level. The software developed has to follow processes which guarantee robustness and performance, and it has to be tested (through the usual quality procedures, and through appropriate benchmarks); this is part of the specific workpackages. Although usability aspects of the components can be subject of individual evaluation outside this workpackage, they will be tested in an integrated way more appropriate to the goals of the project within the framework of the experimental productions. As indicated elsewhere, the detailed procedures for this usability of components will be put in place after the initial phase of prototypes of experimental productions.

Each commercial partner will support the evaluation methodology through the application of its normal commercial evaluation procedures and may at its sole discretion also apply some of the other evaluation methodologies that other partners consider interesting. Evaluation methodologies and results will be shared with all partners.

It is customary to distinguish between *formative* and *summative* evaluation. The former refers to evaluation taking place in the development cycles, assessing the results of one of the cycles to give suggestions for improvements for the next one. The latter refers to evaluation for a released (or to be released) product, and sets to give objective measurements relative to the competing products. SALERO being an RTD project, most evaluation will be formative. Only for several items, which are not identified yet, summative evaluation will be relevant. Moreover, as we have indicated just above on the grounds of the innovations to be introduced by SALERO, our evaluation methodology will be refined and consolidated after the first prototypes cycle.

5 Additional Categorisation of Methods

A valuable perspective on the methods identified above is provided by the framework proposed in the eLearning context by Oliver, M. 1998³. Three criteria are identified: scale, exploration and intervention:

- *Scale* is measure of the number of participants who need to be involved in this type of evaluation.
- *Exploration* refers to a continuum between those methodologies, which provide support for testing (and so dismissing or favouring options) and those, which are about understanding (and so increasing the range of options available). A 1-4 scale is used, with the following suggested meanings:
 1. testing a hypothesis;
 2. considering a range of possibilities;
 3. generating a range of possibilities, focused on a particular topic or issue;
 4. generating possibilities based on an immersion in the context.
- *Intervention* in context distinguishes between methodologies, which are about controlling possible influencing factors from those that are about understanding the effects of context. A 1-4 scale is used, with the following suggested meanings:
 1. Any factors that could influence the intervention are controlled whenever possible.
 2. The group is sampled in some way, but beyond this, factors are not controlled.
 3. An existing group is used, but the evaluation is intrusive.
 4. The evaluation takes place without disturbing the existing context.

When these criteria are applied to the methods to be used in SALERO the results are as follows:

Methods	Exploration	Intervention in context	Scale
1. Cost / benefit analysis	1	2	1
2. Experiment	1	1	30+
3. Observation	2	2	1+
4. User diaries	2	2	1+
5. Questionnaire	2	2	30+
6. System log data	2	4	10+
7. Reviews	3	2	1+
8. Focus group	3	2	4-8
9. 'Open' questionnaire / interview	3	2	10-30
10. Textual data	3	4	10+
11. Case study	4	4	1+
12. Ethnographic study / Contextual design	4	4	1+

Thus it may be seen that the proposed methodologies are evenly spread over the range of the "exploration" scale, and the methodologies are ordered according to their score on this scale. The

³ Oliver, M "Innovation in the Evaluation of Learning Technology". University of North London, 1998. The methods and criteria described here are closely based on the summary of Oliver in the ELT toolkit. The table, which follows, however, includes a different range of methodologies and weightings to those selected by Oliver and ELT toolkit.

“intervention in context” scale is less evenly spread, but there are nevertheless methodologies from both extremes of the scale.

6 Categories of Evaluation Actors, Actions and Areas

6.1 Actors

The level of involvement of the participants in the experimental productions is likely to shift over time, as they will follow a spiral cycle pattern. For example, the initial 18 month period involves experimental production prototypes, and the involvement of final media consumers is likely to be very reduced at this stage; similarly, some of the tools are likely to be at an initial stage, so that the media creators will be required to have a technical competence too, or be heavily supported by a technical person from the research team. On the other hand, as it happens usually actors refer to roles that users can take up, rather than individual people.

The following actors are involved as respondents in the evaluation process

- Professional participants in pre-production
- Digital Content Creators, Developers and other professional participants and operators in production
- Media production supervisors
- Managers
- Audience

And external users of these categories through SALERO User Group, that help validating the results of the project in a more external way.

We should remark that these are broad categories of users, which have to be made more precise for each experimental production. For instance, in the Pre-School one, the specific intended audience are likely to be pre-school children; the scenario has to make more precise the type of content/media intended (e.g. animations, internet games) and the type of content / media will mean different types of professionals involved. (e.g. modellers, animators, rendering technicians, postproduction and audio specialists in case of animation)

6.2 Evaluation Committees

Apart from the evaluation respondents, evaluation needs evaluators who supervise the process and obtain conclusions. SALERO experimental productions are significant pieces of work and each experimental production has an *experimental production evaluation committee* (chaired by the responsible of the experimental production) and an evaluation responsible from each partner taking part in the experimental production.

The overall evaluation is responsibility of an *evaluation committee*, composed by a responsible from each experimental production, and the evaluation coordinator from FBM-UPF.

6.3 Actions

After the methodological discussion, and the actors likely to be involved, let us summarise the main evaluation actions for SALERO experimental productions.

6.3.1 Case studies

A narrative of the experimental productions will be built for evaluation purposes. The narrative will include the scenario definition, the specifications, the development process (from raw material to finished product – which might be a series of prototypes with different degree of finalisation). It should help the

evaluation analysis. The different elements used, the different roles, and the details should be included. The case study description should adopt the ethnographic / contextual design approach. If possible, observation by evaluators of users in the experimental productions will be included.

6.3.2 Cost Benefit Analysis

This will be a key part of the evaluation of experimental productions, and will include the “fitness for use” of the new tools and methodologies of SALERO, and the savings of the approach.

6.3.3 Focus Groups/Workshops

The experimental productions seem to indicate that focus groups of the different types of users involved in experimental productions will be one of the most important methods for their evaluation. Special workshops/focus groups will be organised and held at the end of each experimental production.

6.3.4 Interviews

Semi-structured interviews will be an alternative way of getting enough material for the cost-benefit analysis and the fitness-for-use of SALERO. They will involve the actors described above except for the final users.

6.3.5 Questionnaires

Structured questionnaires will be another alternative for capturing evaluation material, and this will include final users.

6.3.6 Reviews

Internal peer-reviews are an important aspect of SALERO quality control procedures, and reviews on the results of experimental productions, and their evaluation will be another useful tool.

These will be the main methods used for the SALERO integrated evaluation. Other methods we have mentioned such as Experiment, System log data, Textual data generated by users, will not be used in this process although they might be used for specific components.

6.4 Areas and Dimensions

On the basis of SALERO intended innovations, we briefly discuss the key areas which have to be evaluated in the context of media production. After the specifications of the productions are made, the evaluation instruments will have to address them in more detail, in order to provide measurements of success. The main areas will be:

6.4.1 New methodologies

SALERO will provide methodologies for intelligent content, for content re-usability at a new level. It is difficult to assess methodologies and more when they represent introducing novel contents and concepts. The main factors for assessment are probably the “fitness-for-use” and the cost-benefit analysis. It seems that managers and media supervisors will be the main actors involved in this evaluation.

6.4.2 Workflow improvements

The same parameters: “fitness-for-use” and cost-benefit analysis seem to be the ones to measure workflow improvements. While managers and media supervisors will be the main actors involved in this evaluation, professionals involved will be also a key source for the evaluation.

6.4.3 Tools

At the integrated evaluation level we are dealing with it in this workpackage, the main factors for assessing the tools will be their suitability in the experimental productions, their ability to be integrated in

the workflow production. Media supervisors, and professionals involved in the experimental productions will be the main sources for evaluation.

6.4.4 Quality of content produced

SALERO approaches represent a novelty in media production, involving a degree of re-use, intelligent adaptation, and automation of processes. As indicated in the Description of Work, the quality of the results has to be analysed. While the details will be provided after the specifications of the experimental productions are provided, it is likely that the aspect of objects (e.g. colour, integration in the scenes), behaviour of objects (e.g. movements, sounds), and the overall perception of the programmes in some cases, will have to be assessed against the current “manual” alternatives. The main actors in this evaluation will be the professionals involved in production, especially the digital media content creators.

6.4.5 End user satisfaction

While we have just addressed the quality of the media products from the point of view of the content producers, another important source for quality assessment will be the point of view of end users. We have already discussed briefly that end users will probably be different in the wide range of experimental productions. Also, as they will be experimental productions which are not intended for public release, at least at the initial stages, reliable results will be hard to get. However, once the experimental productions are defined, the target users will be defined more precisely and some measures of audience satisfaction will be necessary with some users belonging to the target group.

6.4.6 Transferability

At this early stage of the project, which deals with research, development, and innovation, the main concern is to provide innovative solutions. However, one of the issues of our research is to be able to provide with methods which are transferable, and with tools and new type of content which are interoperable, so that the project gives open solutions which can be used beyond the project participants.

7 Schedule of Evaluation

As indicated at the beginning SALERO follows a type of spiral cycle development. Relating to experimental productions the first stage involves the production of prototypes (and their evaluation) by month 18, which is prepared by the definition of scenarios by month 8, and the specifications by month 12.

The next table provides in a tabular form a schedule for the instruments and outputs for the evaluation of the first experimental productions, which will appear in prototype form.

Action	Date	Content	Main actors	Type of output
Definition of the Evaluation Working Committee	M9	Evaluation representatives of each partner involved in the evaluation plus the evaluation coordinator, who will oversee the whole evaluation of the project	Partners involved in the evaluation of the project	Formal constitution of the committee.
Case studies v0.1	M9	Outline and scenarios of experimental productions	Experimental production managers	Report
Case studies v0.2	M13	V0.1 + specifications	Experimental production managers	Report
Cost benefit evaluation plan	M13	Items to be assessed and actors taking part for each experimental production	Managers	Report
Focus group / workshop planning	M13	Agenda for workshop including questions; identification of participants	Evaluation responsible for each production	Report
questionnaires for interviews with experimental production participants	M13	Model questionnaires for each subgroup Identification of persons / actors to be interviewed	Evaluation responsible for each production	Report and questionnaires
Structured questionnaires for final users	M13	Model questionnaires and identification of target groups	Evaluation responsible for each production	Report and questionnaires
Structured questionnaires for experimental production participants	M13	Model questionnaires and identification of target groups	Evaluation responsible for each production	Report and questionnaires
Update of Evaluation Methodology and plan	M13	Update of Evaluation Methodology in view of specifications, making more precise the actors, their numbers, the parameters to be evaluated for each experimental production and measures.	Evaluation coordinator	Report + annexes – including all above elements for M13
Case studies v1.0	M18	V0.2 + experimental production development	Experimental production managers	Report

Action	Date	Content	Main actors	Type of output
Cost benefit analysis	M18	Items assessed and participants in the assessment for each production	Managers	Report and conclusions
Experimental production workshops, focus groups	M18	Workshops with different sectors for each prototype	Experimental production participants	Workshops summaries. Report and conclusions.
Semi structured interviews	M18	Interviews with participants in experimental productions	Experimental production participants	a) Initial collated responses b) Report with conclusions
Questionnaires with final users	M18	End users feedback	Final users	a) Initial collated responses b) Report with conclusions
Update of Evaluation Methodology	M20	Update of Evaluation Methodology and plans, taking into account the results up to m18	Evaluation coordinator	Report + annexes

Let us remark that the first phase of SALERO (up to month 18) is expected to have prototypes of experimental productions only. At that stage SALERO solutions are likely to be very rough tools and initial methodologies for re-usability and intelligent content. The main evaluation goal of this phase will be to assess the proper orientation of the tools, and the project at this stage should provide more a proof of concept than actual smooth solutions in methodologies and tools.

The update of the evaluation methodology and plans will be much more grounded after the experience of this initial phase. It is likely that for the second phase, much more detail will be needed and provided about components evaluation, not part of the integrated SALERO evaluation. Methods which have not been found useful in this first phase will maybe prove valuable at a later stage.

8 Conclusions

This deliverable has introduced the main methodological aspects of SALERO integrated evaluation and has provided

- a schedule of preparation of instruments and
- a definition of evaluation outputs, for the first 18 months stage of the project.

This will serve as the basis for the refinement and consolidation of the methodology for the remainder of the project.

9 References

- [1] Denzin, N.K. and Lincoln, Y. (eds.). *The SAGE Handbook of Qualitative Research. 3rd edition.* SAGE Publications, 2005.
- [2] Ericsson, K.A. and Simon, H.A. *Protocol Analysis. Verbal reports as data. Revised edition.* The MIT Press, London, England, 1996.
- [3] Morgan, D.L. *Focus groups as qualitative research.* Sage Publications, 1988.
- [4] Nielsen, J. Heuristic evaluation. in Nielsen, J. and Mack, R.L. eds. *Usability Inspection Methods,* John Wiley & Sons, 1994, 25-63.
- [5] Nielsen, J. *Usability Engineering.* Academic Press, Boston, 1993.
- [6] Nielsen, J. and Mack, R.L. (eds.). *Usability Inspection Methods.* John Wiley & Sons, 1994.
- [7] Robson, C. *Real World Research: A Resource for Social Scientists and Practitioners-Researchers.* Blackwell Publishers, 1993.

10 Glossary

Terms used within SALERO project sorted alphabetically.

DoW Description of Work (part of the Contract)

WP Workpackage

Partner Acronyms

AM Activa Multimedia, ES

BLITZ Blitz Games, UK

CINESITE Cinesite Europe Ltd., UK

DIT Dublin Institute of Technology, IE

DTS Digital Theatre Systems, UK

FBM-UPF Fundació Universitat Pompeu Fabra, ES

GVG Grass Valley Germany, DE

JRS JOANNEUM RESEARCH Forschungsgesellschaft mbH, AT

LFUI Leopold-Franzenzs Universtät Innsbruck, AT

PGP Pepper's Ghost Productions Ltd., UK

TAIK Taideteollinen Korkeakoulu, FI

UG University of Glasgow, UK

URL Universitat Ramon Llull, ES