Deliverable D 6.3a

Self-assessment Checklist

Contract number**: FP7-231868 SERA**

Social Engagement with Robots and Agents

The research leading to these results has received funding from the European Community's Seventh Framework Programme (*FP7/2007-2013*) under *grant agreement* n° 231868.

Identification sheet

Project ref. no.	**FP7-231868**
Project acronym	SERA
Status & version	V3 final
Contractual date of delivery	month 3
Actual date of delivery	22-05-09
Deliverable number	D 6.3a
Deliverable title	Self-assessment checklist
Nature	report
Dissemination level	PU
WP contributing to the deliverable	6
WP / Task responsible	OFAI
Editor	Sabine Payr
Editor address	OFAI Freyung 6/6 A-1010 Vienna, Austria sabine.payr@ofai.at
Author(s) (alphabetically)	Payr Sabine
EC Project Officer	Pierre Paul Sondag
Keywords	evaluation, internal review, quality assurance
Abstract (for dissemination)	The report covers the quality assurance procedure adopted for the project. The core process is a "270-degree evaluation" which includes definition of criteria as well as review of results by both self (WP leaders) and other (partners).

Contents

1 Introduction

Task 3 of Workpackage 6 (Management) has been described as follows in the Description of Work (DoW):

> Early on, the indicators of goal achevement and expected results will be further developed into a checklist of measurable progress for each work package and for the project as a whole, and procedures of quality assurance will be defined (e.g. internal reviewers).

The present report covers work done to develop this "checklist". However, the development has not led to the definition of a fixed list of criteria and metrics which have been considered too static and not flexible enough for the project's progress. Instead, it describes the quality assurance procedures that have been implemented to build and refine the checklist dynamically throughout the project, and to control work done in the project against its overall objectives.

Section 2 resumes the objectives and expected results of the work packages, as laid out in the DoW, and breaks down indicators to the level of tasks.

Section 3 introduces the process, i.e. the 270-degree evaluation that will be used as an instrument for internal quality assurance.

Section 4 explains the implementation of the evaluation process.

Section 5 gives an outlook on the means to include external peer review in the evaluation process.

2 Project Objectives and Expected Results

2.1 Project Objectives

To start with, an overview of project objectives and expected outputs, together with the workpackages most concerned with producing this output (from the DoW):

Objective	Expected Output	Work Packages
To generate new and much-needed knowledge about social relationships between users and robots	Video corpus on real-life long-term human-robot relationship; analysis results gained through the application of different (qualitative) and inter-subjectively/theoretically validated methods.	WP 2 WP 1
To make progress in the methodology with which social engagement between robots/agents and humans can be studied, tested and evaluated	Theory and methodology survey and critical review, coordinated publications and presentations targeted at the research community; a blueprint for the architecture.	WP 1
To contribute to the development of theoretical foundations of human-robot/agent social relationships		

To develop, on theoretical grounds and empirical results, a social-emotional reference architecture for interactive robots and agents, and to implement it to a proof-of-concept state.	Architecture specification and its proof-of-concept implementation, as an open access research platform.	WP 3
To develop and test a portable showcase of intertwined service applications in this architecture	Interactive portable showcase system of an assistive robot front-end, intertwining two or more services relevant to the target users.	WP 4
To actively promote scientific debate and networking with the goal of advancing European research in this broad and important domain.	In addition to online resources, presentations and publications: a topical symposium, and a joint peer-reviewed publication (journal special issue).	WP 5

Table 1. Project objectives

2.2 Indicators of goal achievement

The following indicators of goal achievement have been identified in the DoW, and are here related to the tasks concerned. We have distinguished here between "main" and "auxiliary" tasks, insofar as the main task is the one that directly produces the outcome by which goal achievement can be evaluated, while the auxiliary tasks have preparatory and/or supporting roles for the main task which, consequently, depends on their quality and outcome.

Indicators of goal achievement	**Expected results**	**main & auxiliary tasks**
Data collection: Quantity of collected data: number of subjects, duration of recording.	Iteration 1: 3 subjects; iteration 2: 6 subjects of which 3 new; iteration 3: 9 subjects of which 3 new; (= 9 subjects total). Minimum recording duration: 1,5 hours (voluntary) video recording per dataset (= 27 hrs total), 5 hours audio recording per dataset (= 90 hrs total).	Main: task 2.3 Auxiliary: tasks 2.1, 2.2, 2.4, 2.5
Data collection: Technical quality of collected data: rate of intelligible/transcribable data = rate of data usable for analysis.	Minimum: 50 % usable data.	Main: task 2.3
Analysis: Generation of significant and	Tokens for each category are found	Main:

inter-subjectively validated (investigator and method triangulated) analytical categories.	in more than one data set. Categories are confirmed by triangulation.	task 1.3
Theory: Definition of the necessary and sufficient indicators (observable phenomena) of sociability in artifacts.	Operationalizable and testable hypotheses for architecture design and field study	Main: task 1.2 Auxiliary: tasks 1.1, 1.3
Methodology: Critical dimensions of data collection and analysis, credibility, transferability and confirmability of methods.	Theoretically founded coding categories (= hypotheses) for subsequent quantitative analyses (= hypotheses testing).	Main: task 1.2 Auxiliary: task 1.3
Reference architecture: Advance beyond state of the art.	Documented progress wrt. the review of state-of-the-art.	Main: task 3.2 Auxiliary: tas, 3.1
Reference architecture: Implementability & respect of design guidelines (classified as core/additional/optional).	Successful implementation; at least core requirements are all met.	Main: task 3.3
Field study setup: Progress in social engagement as measured by the qualitative criteria developed in the project, plus (probably) quantitative aspects such as e.g. frequency and length of interactions, proportion of user-initiated interactions, success rate of interactions, user satisfaction.	Observable behavioural differences in the collected data from iteration 2 and 3 as compared to iteration 1, from “new” or “old” subjects or both; improvement in sociability as evaluated by the indicators developed in WP 1.	Main: task 2.2 Auxiliary: task 2.1, 2.5
Showcase system: Functionality corresponds to scenario specifications (classified as core/additional/optional).	At least core requirements are met.	Main: task 4.2 Auxiliary: task 4.1
Showcase system: User test results: questionnaires, observation, interviews.	Survey with at least 6 test subjects	Main: task 4.3
Dissemination: Website & repository visits, audience reached, response to the Symposium Call.	Minimum outreach: Hits >100, Downloads of resources > 30, Audience > 150, Submissions > 10	Main: task 5.2 Auxiliary: task 5.1

Table 2. Indicators of goal achievement

These indicators of goal achievement have to be considered preliminary insofar as they have been defined globally and before the project start. Therefore, one of the first tasks of the project was to elaborate on these objectives and metrics, establishing a procedure by which they are specified and assessed throughout the lifetime of the project. Both the objectives and the criteria by which to assess them have to be dynamic: the main concern of SERA are advances in research, and these are, by definition, open to new insights and changes. The following section outlines the procedure with which we want to accomodate this flexibility without, however, falling behind the global initial objectives.

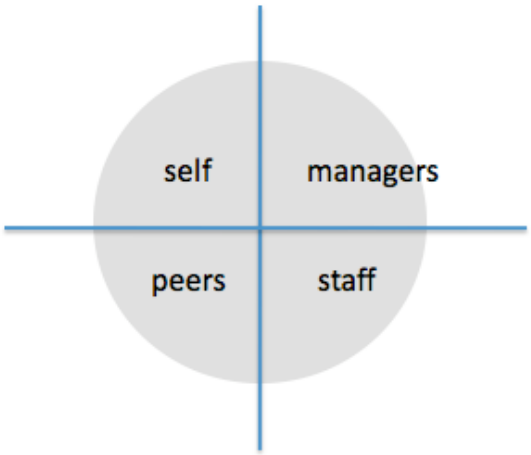
3 Procedure of Quality Assurance

3.1 Rationale

The main idea of the self-assessment in SERA is the definition and refinement of criteria as a collaborative process in which the producers of output, the receivers of output, and the internal reviewers are involved. The procedure that we adopt takes its inspiration from the "360-degree-feedback" instrument.

3.1.1 360-degree feedback

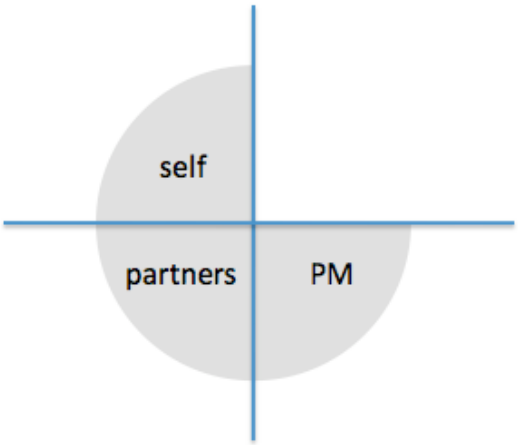
360 degree feedback is a method used in human resources management. Its main goal in these fields is to foster competence development and behavioral change (Blum and Zaugg 2009). The name is derived from the fact that at least three groups of raters are involved, to which self-evaluation is added as the fourth part. Typically, the feedback is given by superiors, peers, and subordinates. The result is compared to the self-assessment by the feedback receiver, a process which improves precision of positioning (ibid.):



3.1.2 270-degree evaluation

The term "270-degree review" is also sometimes used in the Human Resources context, meaning a multi-rater feedback on employees who do not manage staff.

For the project SERA, we have adapted the idea of "270-degree evaluation" with a somewhat different semantics: the three sides of the evaluation being constituted by self (the WP and/or task leader), partners (project beneficiaries not or not centrally involved in the task), and project management.



A further difference is the object of evaluation which, in the case of the project, is a task as defined in the DoW (GA Annex I), or defined part (iteration) where tasks are split up into separate steps (such as in the case of T 1.3 Data collection). A task or iteration is the smallest unit in the project workplan for which efforts and objectives can and have been defined, and thus the most appropriate unit to which to apply quality assurance procedures.

The procedure is also different in that it comprises two steps: the first one occurs at the start of the task, when expectations, criteria and indicators of goal achievement are defined, and the second one at its end or later, when these are reviewed. The intention inherent in this two-step evaluation is to take into account the principle of project quality management whereby "preventive" controlling (or "design control") is an important tool to meet quality requirements (Schelle 2007, p. 203)

What is adopted from the HR model is the idea that multiple sources of evaluation are better than a single judgement, and that these sources include the task leader him/herself. In the following section, we will describe the elements of this 270-degree evaluation for one task.

3.2 Elements of quality assurance

3.2.1 Objectives and indicators

The objectives of the task as they are laid down in the DoW are resumed here, including statements about the outcomes that have been made in describing the task.

Wherever the overall indicators of goal achievement concern the task in question, these are also copied from the DOW and presented.

3.2.2 Task leaders' objectives and criteria of success

At the start of the task, the task leader is required to specify these objectives and to formulate criteria of success. In some cases, job and task descriptions of co-workers can be used in this definition. Through this prcess, we expect to arrive at outcomes and quality criteria that are very close to the task's content and intrinsic to it, instead of using extrinsic, necessarily more standardized and generalized criteria.

The purpose of formulating and specifiying quality criteria early on is to orient and to "keep on track" the execution of the task and to raise awareness of quality issues.

The task leaders have to

- adapt the objectives of the task as laid down in the DoW to possible changes in content (made necessary, for example, by technical, organizational or time constraints, and by results and insights of the tasks feeding into it)
- specify and operationalize them in order to make quality control possible
- identify indicators and criteria of success, together with the estimated date at which they can be determined
- indicate how formal or informal external peer review will be organised (where applicable)

3.2.3 Partners' requirements

Partners not, or not centrally, involved in the task should also formulate their criteria and requirements for the outcome of the task, in quantity and/or quality.

In particular, the workpackage and task leaders at the receiving end of the task in question can, through this instrument, act as the "customers" in formulating their specific needs with regard to the results, which they will have to use to build their work on.

The format of these requirements is open, i.e. no specific questions have to be answered.

3.2.4 Self-assessment

After the end of the task, the task leader is again requested to assess whether and to what degree expectations have been met, and whether the results satisfy the criteria defined beforehand.

This self-assessment can occur in several stages: one takes place at the end of the task in any case. If there are results that are expected with delay, self-assessment will be completed once these results have been achieved. At the latest, a final round of self-assessment will be done at the end of the project.

3.2.5 Internal review

Before a deliverable is submitted or a part of a task is finalized (not all partial tasks end with deliverables, e.g. the data collection and analysis iterations), the appointed internal reviewers will evaluate the deliverable in time for the authors to make improvements and changes.

The internal review results will also be published on the wiki and thus complete the 270-degree evaluation.

Internal reviewers have been appointed at the kick-off meeting on the level of workpackages. As there are hardly any tasks where *not* all partners are involved, the guiding principle was to appoint as reviewers those partners who are *least* involved. To increase objectivity of the internal review, the Partner Representatives are invited to appoint, wherever possible, those persons in their teams who have not or hardly contributed to the task in question.

Workpackage, short name	**Appointed reviewing partner**
Workpackage 1, Theory	UT, USFD
Workpackage 2, Field Study	UDE, OFAI
Workpackage 3, Architecture	UDE, USFD
Workpackage 4, Showcase	UDE, OFAI
Workpackage 5, Dissemination	UT, USFD
Workpackage 6, Management	UT, USFD

Table 3. Internal review assignment

OFAI is somewhat under-represented in this distribution, the reason being that, as coordinator, they will have to review and control all project outcomes anyway.

3.2.6 Results

The evaluation and review process as outlined above are project-internal throughout the lifetime of a task. It will be summarized for distribution e.g. in the deliverables and in the progress reports (see, for example, the partial reproduction of the internal review in the annex of D1.1)

The decision to keep the ongoing evaluation and review process confidential has been taken in order to motivate open discussion and criticism. Whereas to the outside, the consortium will tend to act and appear as a unity, internal discussions and even conflicts are not only possible, but desirable if they take place in a climate of trust and collaboration, to improve the project's overall performance.

3.2.7 Peer Review

In addition to the internal quality assurance process, the project seeks feedback and review from the research community, both formally and informally:

- formal: submission of papers for publication, workshop/session proposals to conferences
- informal: invitation to external researchers to give feedback, make suggestions, exchange results, etc.

This part of the quality assurance plan is closely connected to Workpackage 5, Dissemination: the choice of themes, occasions and audiences for disseminating the project results will have an important influence on the quality of the feedback we can expect to get.

For example, the project has submitted a proposal for a special session at ACII'09. The focus of the proposal is on methodology for data analysis, and should be a panel discussion with prominent experts in the field. The purpose, then, is not simply to present the project, but to make its approach and its research questions tangible - and, hopefully, also to collect valuable input on the field study from the experts and the audience.

External feedback, both formal and informal, will be documented in deliverables and progress reports whenever it becomes available. An overview of both internal and external evaluation will be given in deliverable 6.3b (month 24). However, the usual delays in review processes, e.g. in international scientific journals, could have the consequence that review results become available only after the conclusion of the project.

4 Implementation

At the beginning of each task, the PM requests the WP and task leader to specify, within the framework of the objectives and expected results of the project, detailed expectations for the task.

Project partners are also invited to formulate their expectations for the task. This will be most relevant in those task whose outcomes are critical for ensuing tasks or workpackages, because results are directly needed and used.

Both types of success criteria - self and other generated - will be made accessible to the project partners, but may not be modified.

4.1 Instruments

The following form will be used to collect information on the 270-degree review from task leaders, partners (including reviewers) and peers at the beginning and after the end of the task.

Assessment of Task no.	**Plan (to be completed month .....)**	**Assessment**
Task leader		
Objectives: minimum requirements What is the minimum result that has to be achieved so that this task can be called successful?		
Objectives: optimal outcome What do you plan to achieve beyond the minimum requirements, what would be the optimum outcome?		
Indicators of success What are the criteria by which the difference between failure - minimal outcome - optimum outcome can be defined?		
Measures of success What values do these criteria have to take to qualify the outcome as sufficient/good/optimal?		
External review What kind of formal and/or informal external (peer) review do you plan for this task?		
Partners & internal reviewers		
Objectives and requirements What is the minimum result that has to be achieved so that this task can be called successful, and the project can continue? What outcomes are desirable beyond the minimum?		
Partner:		
Partner:		
Partner:		
Indicators of success: What are the criteria by which the difference between failure -		

minimal outcome - optimum outcome can be defined?		
Partner:		
Partner:		
Partner:		
Measures of success What values do these criteria have to take to qualify the outcome as sufficient/good/optimal?		
Partner:		
Partner:		
Partner:		
External reviewers		
Progress towards project objectives Do the results contribute towards the project's goals as set out in the description of the task?		
Progress beyond state-of-the-art In a larger perspective, do the results of the task represent an advance beyond state-of-the-art?		
Comments & criticism What could/should be corrected, improved, changed, etc.? What should be done to make the work more innovative, relevant, etc.? What should be taken into account in future work?		

Table 4. 270 degree review form

The contents of the forms will be published on the Wiki-like groupware portal (dokuwiki; http://www.splitbrain.org) that is also used as project-internal document server and discussion forum. Partners have been working with this groupware tool since the preparation of the proposal and have become sufficiently familiar with it.

The information is also transferred to a spreadsheet which contains, in addition to the above, the estimated point in time when "delayed" results are expected. For example, if a peer-reviewed publication is put down as a criterion for a successful outcome of a task, this result will usually materialize much later than the task ends. The spreadsheet keeps these dates in evidence and thus allows to check on the results of task later on, e.g. in PMC meetings and in progress

reports. The spreadsheet is also used to store the information published on the wiki at the "cut-off" date for editing, so that it is not possible to change expectations as a task progresses.

4.2 Schedule of review

The planned outcomes and expectations for the task have to be formulated by the task leader at the latest one month after the start of the task in question.

Partners and external reviewers can enter and edit their requirements also later on during the lifetime of the task, because some required results may become evident only later on e.g. through progress in concurrent tasks.

The final assessment of the task coincides with the deadlines for production and delivery of the deliverable, or follows a corresponding timeline for tasks or part of tasks that do not produce independent deliverables. Continuous or near-continuous tasks (e.g. 1.2, 6.2) are additionally evaluated at the occasion of their interim delivering or reporting.

The following table shows the schedule of assessment in detail. Note that in some cases, task leadership may be assigned to a different partner if this is indicated by the division of labour in the workpackage.

Task or iteration	Plan due at start of project month ...	Asessment due until end of project month ...	WP leader
1.1 Literature review	2	5	UDE
1.2 Theory dev.t	6	12, 24*	UDE
1.3 iter. 1 Data analysis	7	8	UDE
1.3 iter. 2	14	15	UDE
1.3 iter. 3	22	23	UDE
2.1 Scenario dev.t	2	5	UDE
2.2 iter. 1 Dev.t&Setup	5	6	USFD
2.2 iter. 2	9	13	USFD
2.2 iter. 3	16	21	USFD
2.3 iter. 1 Data collection	6	7	USFD
2.3 iter. 2	13	14	USFD
2.3 iter. 3	21	22	USFD
2.4 ASR implementation	2	5	USFD
2.5 ASR adaptation	5	13	USFD
3.1 Design guidelines	8	11	UT
3.2 Ref. Architecture	11	14	UT
3.3 Implementation	14	17	UT

4.1 Showcase scenario	14	15	USFD
4.2 Showcase integration	17	24	USFD
4.3 Showcase test	23	24	USFD
5.1 Dissemination plan	2	7	OFAI
5.2 Dissemination activities	6	24	OFAI
6.1 Data Sharing Agr.t	2	3	OFAI
6.2 Coordination	2	13, 24*	OFAI
6.3 Quality Assurance	2	13, 24*	OFAI

Table 5. Schedule of task asessment

4.3 Responsibilities

The Project Manager and leader of Workpackage 6 (Management) is responsible for monitoring the evaluation process, namely by

- storing and making available information (forms, documents, wiki)
- reminding task leaders to publish their objectives, success criteria and metrics in time, as well as their self-assessment
- inviting other partners to state their requirements,
- organising timely internal reviews of deliverables, and
- checking on mid- and long-term results

Issues of quality and quality assurance are put on the agenda of each PMC (every 3 months).

5 Outlook

The dynamic form of quality assurance adopted for the project requires a high degree of collaboration and cooperation from the partners. Additionally, a certain level of trust and openness are required in order to allow for open discussion and criticism.

These conditions contain a certain risk of failure for the quality assurance plan as laid out above. The plan has been approved by the Project Management Committee in its first meeting (31 March 2009), but the PM is aware that constant monitoring and guidance of the process is necessary to make it work.

In every PM meeting, quality assurance issues will be put on the agenda. The milestone for the establishment and functioning of the process has been set as early as Month 6 (2nd PM Meeting). If the expected outcome cannot be met, the procedure will be replaced by more traditional "top-down" quality criteria and methods of measurement.

The final self-assessment results will be collected in deliverable 6.3b (month 24), which will also contain the evaluation of the adopted quality assurance strategy itself.

References

Blum, A., Zaugg, R. J. (2009). 360-Grad-Feedback. Komplexe Arbeitsbeziehungen erfordern differenzierte Feedbacksysteme. In: Thom, N., Zaugg, R. J.: Moderne Personalentwicklung. 3rd ed. Wiesbaden: Gabler. pp. 66-84.

Schelle, H. (2007). Projekte zum Erfolg führen. 5th ed. München: dtv.