PESCaDO

FP7-248594

Personalized Environmental Service Configuration and Delivery Orchestration

SEVENTH FRAMEWORK PROGRAMME

D1.1 Initial Self Assessment Plan

Due date of deliverable: 30.06.2010
Actual submission date: 30.06.2010

Start date of project: 1st January, 2010 Duration: 36 months

Lead contractor for this deliverable: UPF V1

D1.1	Initial Self Assessment Plan
Project Acronym :	PESCaDO
Contract No :	FP7 - 248594
Due Date :	30.06.2010
Reply To:	Leo Wanner leo.wanner@upf.edu
Actual date of delivery:	30.06.2010

Deliverable Identification Sheet

Project ref. no.	FP7-248594
Project acronym	PESCaDO
Project full title	Personalized Environmental Service Configuration and Delivery Orchestration
Security (distribution level)	PU
Contractual date of delivery	Month 6, 30.06.2010
Actual date of delivery	Month 6, 30.06.2010
Deliverable number	D1.1
Deliverable name	Initial Self Assessment Plan
Type	Report
Status & version	Submitted, V1
Number of pages	25
WP / Task responsible	UPF
Other contributors	All partners of the Consortium
Author(s)	Leo Wanner
Internal Reviewer	All partners of the Consortium
EC Project Officer	Manuel Monteiro
Abstract	The purpose of this deliverable is to outline the strategies to be adopted for the assessment of the activities within the WPs and within the individual tasks of the WPs as specified in the Work Plan. The proposed strategies are activity type - and thus task-specific. They range from user evaluation to quantitative measures that assess the performance of some algorithms.

Table of Contents

Executive Summary

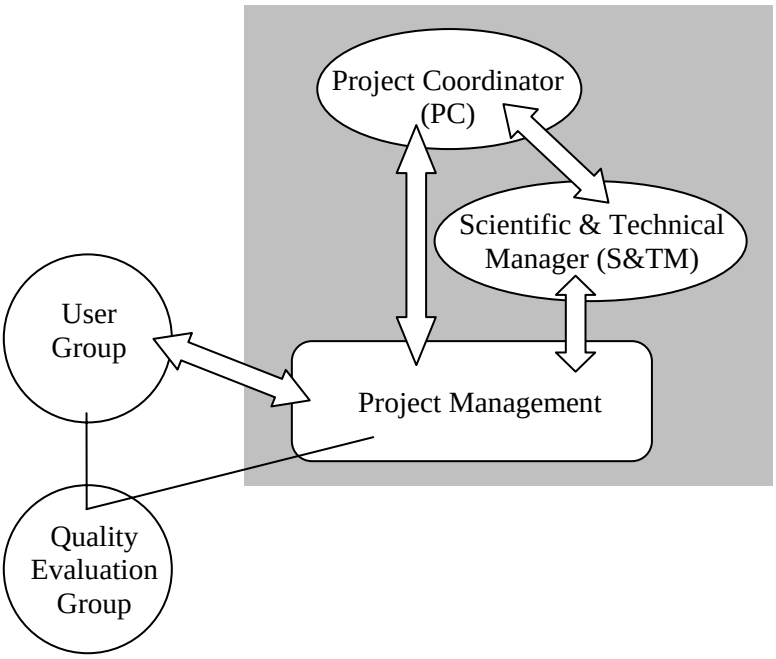
The Initial Self Assessment Plan (ISAP) has the goal to outline the methodology for the surveillance and evaluation of the quality of the performance of the individual tasks as identified in the individual WPs of the Work Plan. This concerns both R&D tasks and the production of deliverables. The surveillance and evaluation is ensured by an efficient management structure that foresees the Quality Evaluation Group (QEG) as a body within the Consortium that is dedicated specifically to this task, and that involves, as needed in each concrete case, expertise from the individual WPs. As assessment instruments, desk studies by experts, qualitative evaluation that draws upon a catalogue of test scenarios, user-targeting questionnaires, and quantitative evaluation in terms of precision, recall, and f-score are foreseen. These instruments aim to evaluate the quality of the work carried out in PESCaDO in terms of its capacity to meet the objectives as formulated in the Work Plan, to contribute to the state-of-the-art and to satisfy the needs of the different types of users targeted by the PESCaDO service. For the R&D oriented WPs, the Initial Assessment Plan uses the assessment measures, indicators and baselines already foreseen for the R&D activities in the Work Plan, Section B1.2.2; the deliverable lists and refines them when appropriate. For the other WPs, the deliverable suggests evaluation strategies and indicators.

1 Introduction

The Initial Self Assessment Plan (ISAP) has the goal to outline the methodology for the surveillance and evaluation of the quality of the performance of the individual tasks as identified in the individual WPs of the Work Plan. This concerns both R&D tasks and the production of deliverables. The surveillance and evaluation is ensured by an efficient management structure that foresees a body within the Consortium that is dedicated specifically to this task and that involves, as needed in each concrete case, expertise from the individual WPs. In what follows, first this management structure is described (Section 2). Then, in Section 3, details of the ISAP for each WP and for each major task within each WP are outlined. This ISAP will be updated and further completed in the course of the project in regular time intervals.

2 The Self Assessment Management Structure

As foreseen in the Work Plan, the management and control of the performance of the self assessment procedures and the evaluation of their outcome will be assumed by the Quality Evaluation Group (QEG), which is interconnected with the Project Management Board and the User Group. The following extract of the corresponding diagram from the Work Plan illustrates the embedding of the QEG into the overall management structure of PESCaDO.



As indicated in the Work Plan, the QEG is composed by a representative from each partner, the PC and the S&TM. When necessary (e.g., when the assessment instruments or the interpretation of the assessment outcome require so), QEG will consult the User Group and WP leaders.

3 Assessment Instruments and Categories

In order to be able to carry out high quality assessment, we need to specify the adequate instruments to be used for the assessment and the categories along which the assessment will be carried out.

3.1 Assessment Instruments

The following four assessment instruments will be applied in PESCaDO:

(i) desk studies by experts;
(ii) qualitative evaluation using a catalogue of test scenarios;
(iii) user-targeting questionnaires;
(iv) quantitative evaluation using baselines and gold standards as references.

During a desk assessment study by an expert, the expert in the field of the component/ result of a task that is to be evaluated assesses the quality drawing upon its personal expertise, the state of the art in the field and the objectives formulated in the Work Plan for this component/task. He/she gives their opinion in writing on the relative and absolute quality and, if appropriate, provides recommendations on the improvement of the quality.

A qualitative evaluation by external or internal evaluators requires the compilation of a catalogue of test scenarios, with each scenario capturing a setting that makes explicit the performance of the tested component in a specific context, and each scenario containing the desired behaviour of the component. This implies that it must be ensured that for all components and all major tasks all potentially relevant contexts are covered by the scenarios used during the evaluation, and that for all of these scenarios the desired performance is specified. In PESCaDO, the task / component responsibles will be responsible for the compilation of the corresponding scenarios; these scenarios will be then validated by the QEG. Once the appropriate test scenarios have been compiled, the qualitative evaluation consists in executing the component in question and comparing its behaviour with the desired behaviour.

User-targeting questionnaires capture the assessment of the quality of a component from the perspective of the different types of users. For each major type of users, a separate questionnaire needs to be worked out. This questionnaire has to contain questions concerning all relevant aspects of the component that come into play when decision support for the users with the corresponding profile is being provided. The questionnaires are also a means to evaluate the satisfaction of the user with the service.

Quantitative evaluation in terms of well-defined measures such as *precision*, *recall*, *f-score*, etc. is the most objective and reliable way to assess the quality of the performance of many tasks of the PESCaDO service – although not of all (consider, e.g., the visual interaction of the service with the user). It presupposes the acquisition of gold standards (i.e., of ideal outputs of the tasks in question) and external baselines against which the actual performance is contrasted.

3.2 Assessment Categories

The assessment in PESCaDO will cover the following three major categories:

(i) the capacity of the results of the task or of a component to meet the objectives as formulated in the Work Plan;
(ii) contribution to the state-of-the-art;
(iii) satisfaction of the user.

For the assessment of the capacity of the evaluated subject matter to meet the objectives of the Work Plan, all four instruments will be used – although the user-targeting questionnaires will play a less central role than the other three.
For the assessment of the contribution to the state-of-the-art by the work done in PESCaDO, expert desk studies and quantitative evaluations (with state-of-the-art approaches serving as baselines) will be primarily used.
To assess the satisfaction of the user, user-targeting questionnaires are the natural evaluation instrument.

4 Self Assessment Plan Broken Down to Tasks in WPs

This section contains the tables for the assessment of the individual tasks within the WPs from the perspective of the objectives they help to achieve.
A separate table is foreseen for each WP. Each table enumerates in its upper part the objectives of the WP, the tasks foreseen to achieve these objectives and the Milestone(s) to which an objective/ a task contribute. In the second part of the table, the evaluation instruments that will be applied to assess the progress and quality of the performance of the tasks are listed. As outlined in Section 3, an instrument can consist in the conduction of user interviews or questionnaires filled out by users, qualitative evaluation against test scenarios, or quantitative evaluation formula (e.g., of precision and recall of an algorithm), etc. – depending on the nature of the task. In its lower part, a table contains the indicators and baselines used for the individual instruments and their instantiation by the time of the compilation of the deliverable.
For the R&D oriented WPs, the Initial Assessment Plan will start from the assessment measures, indicators and baselines foreseen for the R&D activities in the Work Plan, Section B1.2.2 (repeated and refined below) and adapt them as needed in the course of the project. Any major modification of an assessment strategy is to be approved by the QEG. For the remaining WPs, the initial strategies and the indicators are as presented below.
From the perspective of the assessment category “contribution to the state-of-the-art” (see Subsection 3.2), the minimal expectation towards all R&D techniques is that they perform slightly better than state-of-the-art techniques; the maximal expectation is that they perform significantly better than the state-of-the-art techniques. The precise definition of “slightly” and “significantly” is to be determined in the course of the evaluations of each technique in consensus with experts in the corresponding field.

4.1 WP1: Management

WP task Assessment			
WP	1	Date:	30.06.2010
resp. Partner	UPF	Signed:	L. Wanner

1. WP Objectives			
#	*Objective*	*Task*	*Milestone*
1	Ensure that the project is carried out in accordance with the given time and budget parameters.	T1.1	MS1 – MS5
2	Ensure the high quality of the research in the project.	T1.2	MS1 – MS5
3	Ensure the high quality of the SW being developed in the project.	T1.3	MS1 – MS5
4	Ensure proper administration and communication with the Commission	T1.4	MS1 – MS5

2. Evaluation strategy	
#	*Evaluation strategy description*
1	• Request of internal periodic 6thmontly activity and expenditure reports from each partner. • Assessment of the completion (in %) of each task/activity by the time foreseen in the Work Plan and the PESCaDO's Roadmap (D8.2), based on the information obtained from the 6monthly internal activity reports and the interviews with the responsibles of the task/activity. • Assessment of the budget figures as reported by the partners in the 6monthly internal expenditure reports in the light of overall budget figures of PESCaDO and their extrapolation to the total duration of the project.
2	Assessment of the research progress within each WP as documented in the internal periodic reports, deliverables and as presented by the responsibles of the tasks/activities at WP or plenary meetings, according to the metrics or criteria established within the WP in question.
3	Assessment of the SW development progress within each WP as documented in the internal periodic reports, deliverables and as presented by the responsibles of the tasks/activities at WP or plenary meetings, according to the metrics or criteria established within the WP in question.
4	• Assessment of the successful and timely completion of the requests by the partners of Consortium and the Commission. • Assessment of the timely submission of the deliverables. • Assessment of the successful management of the meetings.

3. Indicators		
#	*Highest expectation*	*Lowest expectation*
1	(a) 100% completion of all tasks / activities by the foreseen deadline; (b) expenditure of funds proportionally to the lifetime of the project.	(a) 75% completion of all tasks / activities by the foreseen deadline; (b) justified expenditure of funds, in accordance with the EC regulations.
2	100% fulfilment of the criteria / highest metric figures established for a given task / activity.	meeting the minimal requirements / performance figures established for a given task / activity.
3	100% fulfilment of the criteria / highest metric figures established for a given task / activity.	meeting the minimal requirements / performance figures established for a given task / activity.
4	(a) completion of all requests by the	(a) completion of all requests by the partners

	partners of the Consortium and the Commission within 7 working days to the satisfaction of the inquirer; (b) submission of the deliverables by the deadline established in the Work Plan; (c) exhaustive treatment of all topics foreseen in the agenda of a meeting.	of the Consortium and the Commission within 14 working days to the satisfaction of the inquirer; (b) submission of the deliverables within the contractually fixed 45 days after the deadline established in the Work Plan; (c) treatment of all topics foreseen in the agenda of a meeting to an extent that allows for the continuation and successful completion of the topics over distance.

While the achievement of objectives 1–3 cannot yet be assessed by the submission deadline of this deliverable (Month 6), the objective 4 has been fully achieved to the highest expectations:

(i) the requests (first of all of the Commission, concerning the change of the Coordinator from BM to UPF and the change of the name of YTV to HSY) have been answered promptly;

(ii) the kick-off meeting in Barcelona and the WP8 meeting in Karlsruhe were to the full satisfaction; and

(iii) all deliverables have been submitted on time, either before the due time or on the day of the deadline.

4.2 WP2: Environmental Service Node Discovery

WP task Assessment			
WP	2	Date:	30.06.2010
resp. Partner	CERTH	Signed:	S. Vrochidis

1. WP Objectives			
#	*Objective*	*Task*	*Milestone*
1	Setting up the search infrastructure in PESCaDO	T2.1	MS2-MS5
2	Identify the search-relevant specifics of the metadata and content codification in web pages containing environmental material	T2.2	MS2-MS5
3	Develop query formulation and expansion techniques: (i) keyword spice oriented querying; (ii) inference-based query compilation / expansion; (iii) user interaction based querying	T2.3	MS2-MS5
4	Develop a functional index repository and indexing procedure: (i) define an environmental node profile record; (ii) develop a node indexing procedure.	T2.4, T2.6	MS2-MS5
5	Develop / adapt an existing keyword-based text and metadata search in web-based sources	T2.5.1	MS2-MS5
6	Develop semantic search in web-based sources	T2.5.2	MS2-MS5
7	Develop language recognition techniques	T2.5.3	MS2-MS5

2. Evaluation strategy	
#	*Evaluation strategy description*
1	Hands-on test of the relevant functionality following a test protocol defined by CERTH and validated by QEG
2	Quantitative evaluation (e.g., in terms of precision and recall) of the performance of the metadata and keyword-based search ran with a dataset that contain a representative variation of the identified parameters against a gold standard.
3	• Quantitative evaluation against a gold standard and comparison in terms of precision-recall with cases in which query expansion is not performed in order to identify improvement. • User satisfaction interviews / questionnaires for the interaction part.
4	• Quantitative evaluation of the metadata and keyword-based search ran with the compiled index against a gold standard. • Time performance of the indexing procedure • Time required for the node retrieval from the repository
5	• Quantitative evaluation of the performance of the search engine (in terms of its precision and recall figures) against a gold standard • Time / memory benchmarking • User satisfaction interviews / questionnaires
6	• Quantitative evaluation of the performance of the search engine (in terms of its precision and recall figures) against a gold standard • Time / memory benchmarking • User satisfaction interviews / questionnaires
7	• Quantitative evaluation (in terms of its precision and recall figures) against a gold standard.

3. Indicators		
#	*Highest expectation*	*Lowest expectation*
1	The search infrastructure should operate	The search infrastructure should operate

	flawlessly, rapidly, effectively and communicate with the other modules based on the predefined protocols.	flawlessly and cooperate at least with the modules that there is dependency.
2	Achievement of 95% of identification (precision and recall) of the search-relevant metadata and content existing in a gold standard document selection.	Precision and recall are close to 75%, thus relevant information is not identified in all documents.
3	Improvement of recall by 10% as precision would remain at least stable compared to the baseline system when query formulation and expansion techniques are applied.	Improvement of recall by 5% compared to the baseline system when query formulation and expansion techniques are applied.
4	Develop an indexing structure that traverses rapidly the structure, enables fast deletion and insertion of data into the structure and is memory efficient.	Develop an indexing structure that focuses only on traversing rapidly the structure.
5	Achievement of 85% retrieval (precision and recall) of the web sources results based on the gold standard through metadata and keyword based techniques. Fast retrieval is also required (some seconds) and low memory usage.	Precision and recall are close to 70% and the system responds slower (some minutes).
6	Achievement of 85% retrieval (precision and recall) of the web sources results based on the gold standard through semantic search. Fast retrieval is also required (some seconds) and low memory usage.	Precision and recall are close to 70% and the system responds slower (some minutes).
7	Achievement of 95% precision of identifying a documents language for all languages found.	Precision of identifying a documents language depends on the language and the minimum acceptable percents are 80% for English documents, 85% for Finnish and 80% for Swedish.

4.3 WP3: Orchestration of Environmental Services

WP task Assessment			
WP	3	Date:	30.06.2010
resp. Partner	FMI	Signed:	A. Karppinen

1. WP Objectives

#	*Objective*	*Task*	*Milestone*
1	Design and realization of a service infrastructure for the interconnection of environmental services, drawing upon data and semantic protocol standards.	T3.1	MS2-MS5
2	Derivation of imprecision metrics for the determination of uncertainty of the output of services originating from lack of trustworthiness, erroneous data, data gaps, etc.	T3.2.1	MS2-MS5
3	Deduction of a parameter set for confidence metrics	T3.2.2	MS2-MS5
4	Derivation of confidence metrics for the assessment of environmental nodes	T3.2.2	MS2-MS5
5	Development of a matchmaking strategy for the selection of the best matching nodes to form a node configuration.	T3.3	MS2-MS5
6	Development of techniques for fusion of information proceeding from different environmental nodes	T3.4	MS2-MS5

2. Evaluation strategy

#	*Evaluation strategy description*
1	Hands-on qualitative evaluation of the relevant functionality following a test protocol defined by IOSB and validated by QEG
2	• Quantitative evaluation against a gold standard • Qualitative user-centered (questionnaire-based) evaluation
3	Quantitative and qualitative evaluation of the (i) representativeness and (ii) distinctiveness of the individual parameters.
4	• Quantitative evaluation against a gold standard • Qualitative user-centered evaluation based on a questionnaire
5	• Quantitative performance evaluation against gold standard configurations • Time and memory-oriented benchmark evaluation • Qualitative user-centered evaluation based on a questionnaire
6	• Quantitative evaluation against a gold standard

3. Indicators

#	*Highest expectation*	*Lowest expectation*
1	100% achievement of the functionality specified in the test protocol	75% achievement of the functionality specified in the test protocol
2	≥90% accuracy against a gold standard; 90%-100% satisfaction of the user (according to the filled-out questionnaires) when ordering a set of environmental nodes with respect to their imprecision	75% accuracy against a gold standard; 75% satisfaction of the user (according to the filled-out questionnaires) when ordering a set of environmental nodes with respect to their imprecision
3	Quantitative parameter evaluation: each parameter is reflected in ≥95% of the test configurations; ≤ 5% of the parameters are not 100% distinctive Qualitative evaluation: in 100% of the test configurations, the removal of any	Quantitative parameter evaluation: each parameter is reflected in at least 75% of the test configurations; ≤ 15% of the parameters are not 100% distinctive. Qualitative evaluation: in 90% of the test configurations, the removal of any parameter

	parameter leads to a decrease in accuracy of the metrics.	leads to a decrease in accuracy of the metrics.
4	≥ 90% precision in the relative ordering of a list of environmental nodes according to their confidence; 90%-100% satisfaction of the user (according to the filled-out questionnaires) with respect to the assignment of confidence ratings to environmental nodes	75% precision in the relative ordering of a list of environmental nodes according to their confidence; 75% satisfaction of the user (according to the filled-out questionnaires) with respect to the assignment of confidence ratings to environmental nodes
5	≥ 90% accuracy against a gold standard of nodes and their configurations; minimal time and memory consumption (benchmarks to be determined in the course of the project); 90%-100% satisfaction of the user (according to the filled-out questionnaires) with the configurations offered by the system	75% accuracy against a gold standard of nodes and their configurations; acceptable time and memory consumption (such that the overall satisfaction of the user is not influenced negatively by long answer times); 75% satisfaction of the user (according to the filled-out questionnaires) with the configurations offered by the system
6	≥ 90% accuracy against a gold standard	75% accuracy against a gold standard

4.4 WP4: Environmental Ontology Construction and Population

WP task Assessment			
WP	4	Date:	30.06.2010
resp. Partner	FBK	Signed:	E. Pianta

1. WP Objectives			
#	*Objective*	*Task*	*Milestone*
1	Provision of an overview of available environmental ontologies	T4.1	MS1-MS5
2	Collection of sufficiently large and rich environmental material corpora	T4.2	MS2-MS5
3	Availability of an operational infrastructure for corpus processing	T4.3	MS2-MS5
4	Development of techniques for automatic alignment and extension of environmental ontologies	T4.4	MS2-MS5
5	Development of techniques for content distillation from multilingual textual material and its subsequent integration into ontologies	T4.5	MS2-MS5

2. Evaluation strategy	
#	*Evaluation strategy description*
1	Expert desk evaluation
2	Expert desk evaluation
3	Qualitative functionality evaluation following a test protocol proposed by FBK and validated by QEG
4	• Quantitative evaluation against a gold standard • Qualitative evaluation based on test protocols proposed by FBK and validated by QEG
5	• Quantitative evaluation against a gold standard • Qualitative evaluation based on test protocols proposed by FBK and validated by QEG

3. Indicators		
#	*Highest expectation*	*Lowest expectation*
1	Evaluating experts attest that all available ontologies have been identified and adequately described	Evaluating experts attest that all major and most relevant ontologied have been identified and adequately described.
2	The collected corpora for all three languages (Finnish, English and Swedish) are sufficiently large and diverse to serve all needs in PESCaDO without that further corpora must be collected in the course of the project	The collected corpora for the three languages are sufficiently large and diverge to realize the first versions of the corpus-based techniques in PESCaDO, although for the advanced versions these corpora must be extended.
3	The qualitative evaluation of the available corpus processing tools (both in FBK and publicly accessible) perform the required tasks in domain-specific environments 100% in accordance with the test protocol set up by FBK for the three project languages.	The infrastructure tools perform 100% for English and 50-75% for Finnish and Swedish (due to the reduced availability of NLP technologies for these two languages).
4	Quantitative evaluation: ≥ 90% accuracy against a gold standard	Quantitative evaluation: 75% accuracy against a gold standard

	Qualitative evaluation: The ontology extension techniques fill in the gaps of selected test fragments of existing ontologies and add all relevant concepts required in the project in 90-100% of the test cases.	Qualitative evaluation: The ontology extension techniques fill in the gaps of selected test fragments of existing ontologies and add all relevant concepts required in the project in at least 70% of the test cases.
5	Quantitative evaluation: ≥ 90% accuracy against a gold standard Qualitative evaluation: All information needed is retrieved to 100%.	Quantitative evaluation: 75% accuracy against a gold standard for English and 65% for Finnish and Swedish. Qualitative evaluation: The information needed is retrieved to 75% for English and to 65% for Finnish and Swedish.

4.5 WP5: Reasoning and User-Oriented Decision Support

WP task Assessment			
WP	5	Date:	30.06.2010
resp. Partner	FBK	Signed:	L. Serafini

1. WP Objectives			
#	*Objective*	*Task*	*Milestone*
1	Availability of a problem description language	T5.1	MS2-MS5
2	Development of techniques for user-assisted ontology-based reasoning	T5.2	MS2-MS5
3	Development of techniques for user- and problem-driven content selection	T5.3	MS2-MS5

2. Evaluation strategy	
#	*Evaluation strategy description*
1	Qualitative evaluation of the: • expressiveness of PDL in covering the relevant user problems relative to the use cases considered; • user satisfaction in easiness of formulation of the request using the PDL
2	• Time and Memory benchmarks • Supported expressiveness of the reasoning task
3	• Quantitative and qualitative expert user evaluation on the performance of the content selection techniques upon a set of representative user requests;

3. Indicators		
#	*Highest expectation*	*Lowest expectation*
1	• 100% coverage of the user relevant problems considered in the pilot use cases, and evidence of the ease adaptability of the PDL to cover more general use cases and applications not considered in PESCaDO. • Maximum (100%) user satisfaction in easiness of formulation of the request using the PDL (according to performed tests)	• 75% coverage of the relevant and most popular problems considered in the pilot use cases, mostly “occasional” problems treated as exceptions. • Average user satisfaction in easiness of formulation of the request using the PDL.
2	• Execution time and resources usage comparable with the 90% of the sum of the execution times and resources usage of state of the art reasoners on the single tasks. • OWL 2 reasoning support	• Execution time and resources usage comparable with the sum of the execution times and resources usage of state of the art reasoners on the single reasoning tasks. • RDFS / OWL 1 Lite reasoning support
3	• 95-100% appropriateness and completeness of the content selected by the system upon a user request	• 75% appropriateness and 75% completeness of the content selected by the system upon a user request

4.6 WP6: Visualization and User Interaction Techniques

WP task Assessment			
WP	6	Date:	30.06.2010
resp. Partner	USTUTT	Signed:	T. Ertl

1. WP Objectives

#	*Objective*	*Task*	*Milestone*
1	Development of interactive techniques for query expansion and user feedback realization in the context of environmental node search	T6.2	MS2-MS5
2	Development of visual analytics techniques for uncertainty metrics determination	T6.1	MS2-MS5
3	Development of interactive techniques for integration of the user into the process of service orchestration.	T6.3	MS2-MS5
4	Development of a graphical PDL editor and interactive techniques for user feedback in the context of decision support.	T6.4	MS2-MS5

2. Evaluation strategy

#	*Evaluation strategy description*
4	• Quantitative: o Percentage of query types that are supported by the visual editor

all	• Quantitative:
	  o Responsiveness and runtime complexity
	  o User satisfaction measured on Likert scale by different criteria
	  o Percentage of reached research indicators (see Description of Work)
	• Qualitative: Evaluation of usability with users of different user groups against a test protocol. The protocol will be worked out by USTUTT in collaboration with FMI and validated by the QEG.

3. Indicators

#	*Highest expectation*	*Lowest expectation*
4	• 100% support for the available query types	• Support for the query types relevant for the demonstration use cases

All	• Quantitative:
	  o Interactive response times without noticeable delay
	  o 90-100% mean user satisfaction level
	  o Reached ≥ 90% of the stated research indicators
	• Qualitative
	  o Developed techniques are an effective and efficient means for fulfilling the user tasks; 100% user satisfaction according to the filled out questionnaires
		• Quantitative:
		  o Interactive response times (possibly with a noticeable delay), considering only the visualization module without the time needed by underlying data providers
		  o At least 70% mean user satisfaction
		  o Reached at least 70% of the stated research indicators
		• Qualitative:
		  o Developed techniques are an effective means for fulfilling the user tasks; at least 75% user satisfaction according to the filled out questionnaires.

4.7 WP7: Multilingual Environmental Information Generation

WP task Assessment			
WP	7	Date:	30.06.2010
resp. Partner	UPF	Signed:	L. Wanner

1. WP Objectives			
#	*Objective*	*Task*	*Milestone*
1	Overview of linguistic constructions used in environmental material in Finnish, English and Swedish that is to be generated in PESCaDO	T7.1	MS1-MS5
2	Development of content selection techniques driven by discourse coherence, user profile and language criteria.	T7.2	MS2-MS5
3	Development of schema-based discourse planning techniques	T7.3	MS2-MS5
4	Development of ML-based discourse planning techniques.	T7.3	MS2-MS5
5.	Development of mode selection techniques	T7.4	MS2-MS5
6	Development of linguistic resources for rule-based sentence generation of Finnish, English and Swedish	T7.5	MS2-MS5
7	Improvement of the rule-based generator available at UPF	T7.5	MS2-MS5
8	Annotation of sufficiently large and diverse corpora for ML-based sentence generation	T7.5	MS2-MS5
9	Further development of the ML-based generator available at UPF	T7.5	MS2-MS5

2. Evaluation strategy	
#	*Evaluation strategy description*
1	Qualitative desk evaluation by an expert of the completeness of the study
2	• Qualitative evaluation against a test protocol. The protocol will be worked out by UPF in collaboration with FBK and HSY and validated by the QEG. • Quantitative performance evaluation against a gold standard. • Time and memory benchmarks
3	Qualitative evaluation against a test protocol. The protocol will be worked out by UPF in collaboration with HSY and validated by the QEG.
4	• Qualitative evaluation against a test protocol. The protocol will be worked out by UPF in collaboration with FBK and HSY and validated by the QEG. • Quantitative performance evaluation against a gold standard. • Time and memory benchmarks
5	• Qualitative evaluation against a test protocol. The protocol will be worked out by UPF in collaboration with HSY and validated by the QEG. • Quantitative performance evaluation against a gold standard
6	Validated in the course of evaluation 7.
7	• Qualitative evaluation against a test protocol. The protocol will be worked out by UPF in collaboration with FBK and HSY and validated by the QEG. • Time and memory benchmarks
8	• Inter-annotator agreement • Quantitative annotation quality evaluation against a gold standard.
9	• Qualitative evaluation against a test protocol. The protocol will be worked out by UPF in collaboration with FBK and HSY and validated by the QEG. • Quantitative performance evaluation against a gold standard. • Time and memory benchmarks

3. Indicators

#	*Highest expectation*	*Lowest expectation*
1	Evaluating experts attest that the most prominent linguistic constructions used in Finnish, English and Swedish in environmental discourse have been identified in sufficient detail	Evaluating experts attest that the most prominent linguistic constructions used in Finnish, English and Swedish in environmental discourse have been identified in sufficient detail to start the work in WP7, but that further studies are necessary to complement the identified constructions.
2	Qualitative evaluation: 95-100% appropriateness of the selected content, according to the test protocol; quantitative evaluation: ≥ 90% accuracy against a gold standard; time and memory benchmarking: optimal performance according to the benchmarks worked in the course of the project.	Qualitative evaluation: at least 75% appropriateness of the selected content, according to the test protocol; quantitative evaluation: about 75% accuracy against a gold standard; time and memory benchmarking: satisfying performance according to the benchmarks (i.e., performance that does not cause noticeable delays in the overall system).
3	95-100% appropriateness of the discourse structure and of the ordering between the discourse elements (according to the test scenarios)	75% appropriateness of the discourse structure and of the ordering between the discourse elements
4	Qualitative evaluation: 90% appropriateness of the resulting discourse structures, according to the test protocol; quantitative evaluation: ≥ 90% accuracy against a gold standard; time and memory benchmarking: optimal performance according to the benchmarks worked in the course of the project.	Qualitative evaluation: 75% appropriateness of the resulting discourse structures; quantitative evaluation: about 75% accuracy against a gold standard; time and memory benchmarking: satisfying performance according to the benchmarks (i.e., performance that does not cause noticeable delays in the overall system).
5	Qualitative evaluation: 95% appropriateness of the selected modi, according to the test protocol; quantitative evaluation: ≥ 90% accuracy against a gold standard	Qualitative evaluation: 75% appropriateness of the selected modi, according to the test protocol; quantitative evaluation: 75% accuracy against a gold standard
6/7	Qualitative evaluation: the generator is able to produce 100% of the samples of linguistic constructions listed in the test protocol; time and memory benchmarks: optimal performance according to the benchmarks worked in the course of the project	Qualitative evaluation: the generator is able to produce at least 75% of the samples of linguistic constructions listed in the test protocol; time and memory benchmarks: satisfying performance according to the benchmarks.
8	High inter-annotator agreement; quantitative evaluation: ≥ 95% annotation accuracy against a gold standard	Medium inter-annotator agreement; quantitative evaluation: 80% annotation accuracy against a gold standard
9	Quantitative evaluation: ≥ 80% accuracy against a gold standard; qualitative evaluation: the generator is able to produce 100% of the samples of linguistic constructions listed in the test protocol; optimal performance according to time and memory benchmarks	Quantitative evaluation at least 60% accuracy against a gold standard; qualitative evaluation: the generator is able to produce 75% of the samples of linguistic constructions listed in the test protocol; time and memory benchmarks: satisfying performance according to the time and memory benchmarks

4.8 WP8: System Development

WP task Assessment			
WP	8	Date:	30.06.2010
resp. Partner	IOSB	Signed:	T. Usländer

1. WP Objectives			
#	*Objective*	*Task*	*Milestone*
1	Roadmap for the development of the PESCaDO platform	T8.1	MS1
2	User profile typology	T8.2	MS2-MS5
3	Specification of the PESCaDO Use cases	T8.3	MS2-MS5
4	Specification of the PESCaDO architecture and set up of an operational infrastructure	T8.4	MS2
5	Development of the first prototype	T8.5	MS3
6	Development of the second prototype	T8.6	MS4
7	Development of the final PESCaDO showcase	T8.7	MS5

2. Evaluation strategy	
#	*Evaluation strategy description*
1	Qualitative desk evaluation by an expert
2	Qualitative desk evaluation by an expert of the completeness and adequacy of the typology
3	Qualitative evaluation against a functionality coverage protocol. The protocol will be worked out by IOSB in collaboration with HSY and validated by the QEG.
4	Qualitative evaluation against a test protocol. The protocol will be worked out by IOSB in collaboration with HSY and validated by the QEG.
5	• Qualitative evaluation against a test protocol. The protocol will be worked out by IOSB and validated by the QEG. • Evaluation of the individual components following the evaluation procedures in the individual WPs.
6	• Qualitative evaluation against a test protocol. The protocol will be worked out by IOSB and validated by the QEG. • Evaluation of the individual components following the evaluation procedures in the individual WPs.
7	• Qualitative evaluation against a test protocol. The protocol will be worked out by IOSB and validated by the QEG. • Evaluation of the individual components following the evaluation procedures in the individual WPs.

3. Indicators		
#	*Highest expectation*	*Lowest expectation*
1	Sticking 100% to the work package descriptions and the timeline estimated in the roadmap.	Sticking 80% to the work package descriptions and the timeline estimated in the roadmap.
2	100% typology coverage.	75% typology coverage.
3	100% query satisfaction of one or more test users for each of the identified user types.	75% query satisfaction of at least one test user for each of the identified user types.
4	Working demonstration of Scenario 1 with dummy (mock-up) implementations of the services.	Identified and installed infrastructure components like map-server, web-server, knowledge base, etc.
5	100% fulfilment of the test protocol. Full implementation of scenario 1 from	75% fulfilment of the test protocol. Implementation of scenario 1 from UC1 with

	UC1.	at least 80% real services (some dummy implementations left).
6	100% fulfilment of the test protocol. Implementations of scenarios which cover the full UC1 and aspects from UC2.2.	75% fulfilment of the test protocol. Implementations of scenarios which cover the full UC1.
7	100% fulfilment of the test protocol. Implementations of scenarios which cover the full UC1, UC2.1 and UC2.2.	75% fulfilment of the test protocol. Implementations of scenarios which cover the full UC1 and UC2.2.

4.9 WP9: Assessment and Evaluation

WP task Assessment			
WP	9	Date:	30.06.2010
resp. Partner	HSY	Signed:	T. Koskentalo

1. WP Objectives			
#	*Objective*	*Task*	*Milestone*
1	Development of research and system quality assurance guidelines	T9.1	MS1
2	Monitoring and assuring the achievement of project's objectives	T9.2	MS1-MS5
3	Development of metrics for the evaluation of the theoretical solutions as well as for the scalability and performance evaluation of the prototypes and the final showcase	T9.3	MS1-MS5
4	Evaluation of the project's results	T9.4	MS1-MS5

2. Evaluation strategy	
#	*Evaluation strategy description*
1	Qualitative desk evaluation by an expert
2	• Request of internal biannual progress reports • Desk evaluation of the progress reports and, if necessary, suggestion of remedial actions
3	Verification of the availability of the metrics for all objectives outlined in this deliverable for WP1-WP10; their adequacy is verified in the corresponding WPs
4	Verification of the availability of the evaluation reports for all objectives outlined in this deliverable for WP1-WP10 Qualitative desk evaluation of the evaluation reports by an expert

3. Indicators		
#	*Highest expectation*	*Lowest expectation*
1	Evaluating experts attest the full adequacy of the guidelines	Evaluating experts attest the adequacy of the guidelines, with minor extensions or corrections
2	The biannual internal progress reports are delivered on time; the evaluation of the progress reports reveals a 100% conformity with the Work Plan and its timeline; no remedial actions are necessary for any of the WPs	The biannual internal progress reports are delivered at most 14 days after the agreed deadline; the evaluation of the progress reports reveals the conformity with the Work Plan and its timeline, with minor delays or deviations; remedial actions suggested to account for these delays/deviations are carried out according to the agreed timetable
3	The metrics are available at least 1 month ahead of critical time points such as the Milestones at which the corresponding techniques have to be operational.	The metrics are available at least 14 days ahead of critical time points such as the Milestones at which the corresponding techniques have to be operational.
4	The evaluation reports are available on time; the evaluating experts attest the full adequacy of the evaluation reports	The evaluation reports are available at the latest 14 days after the deadline; the evaluating experts attest the adequacy of the evaluation reports; with possible suggestions of minor extensions or corrections

4.10 WP10: Dissemination, Demonstration and Exploitation

WP task Assessment			
WP	10	Date:	30.06.2010
resp. Partner	UPF	Signed:	L. Wanner

1. WP Objectives			
#	*Objective*	*Task*	*Milestone*
1	Ensure publication and dissemination of the project results	T10.1	MS1-MS5
2	Ensure market awareness and technology watch	T10.2	MS1-MS5
3	Demonstration of the intermediate and final results of the project	T10.3-4	MS1-MS5
4	Development of a business model	T10.5	MS1-MS5
5	Impact analysis and impact intensification	T10.6	MS1-MS5
6	Collaboration with other projects	T10.7	MS1-MS5

2. Evaluation strategy	
#	*Evaluation strategy description*
1	• Assessment of the number of publications and their distribution across the publication types (journals, conferences, workshops) • Ranking and impact of the journals and conferences in which the results have been published • Assessment of the significance of the events (fairs and other types of gatherings) at which the results have been presented
2	Qualitative desk assessment of the completeness and up-to-dateness of the monitoring protocols
3	Qualitative desk assessment of the protocols of the demonstration of the results
4	Qualitative desk evaluation by an expert
5	Qualitative desk evaluation by an expert Number of citations of PESCaDO publications Number and kind of established contacts
6	Assessment of the statistics on the relevant projects in the field and the established collaborations

3. Indicators		
#	*Highest expectation*	*Lowest expectation*
1	The work on each major task in the R&D WPs leads to 1 journal publication and 5 conference publications; the journal and the conferences belong to the high impact publication media in the field; the events at which the results are presented, belong to the top events in the field	The work on each major task in the R&D WPs leads to 5 conference publications; the conferences belong to the top conferences in the field; the events at which the results are presented, belong to the top events in the field
2	Evaluating experts attest the full adequacy of the market and technology monitoring activities documented in the monitoring protocols	Evaluating experts attest the adequacy of the market and technology monitoring activities documented in the monitoring protocols, suggesting minor extensions or correction of the strategy
3	Evaluating experts attest the full adequacy of the demonstration	Evaluating experts attest the adequacy of the demonstration procedures for presenting all

	procedures for presenting all significant technologies developed in PESCaDO by the given time point	significant technologies developed in PESCaDO by the given time point, suggesting minor modifications or extensions
4	Evaluating experts attest the full adequacy of business model	Evaluating experts attest the adequacy of business model, suggesting minor modifications or extensions
5	Evaluating experts attest a high impact of the results of PESCaDO and approve the impact intensification strategy; the number of citations of PESCaDO publications is in-line with the number of citations of publications of other FP7 projects in the field; contacts are established to all major players in the field identified during the market and technology watch activities.	Evaluating experts attest a high impact of the results of PESCaDO and approve the impact intensification strategy, suggesting minor modifications of the latter; the number of citations of PESCaDO publications is in-line with the number of citations of publications of other FP7 projects in the field; contacts are established to the major players in areas that are covered in PESCaDO.
6	Contacts have been established to all relevant projects and an R&D exchange has been agreed upon	Contacts have been established to all relevant projects and collaboration is being sought.

5 Conclusions

The Initial Self Assessment Plan (ISAP) outlined in the sections above presents the categories along which each major task within the individual WPs of the Work Plan will be evaluated, the instruments which will be applied for the evaluation and the indicators that will be used during the evaluation. The WP-specific ISAPs as contained in Section 4 will allow for a close monitoring of the research and development carried out within PESCaDO and thus ensure high quality intermediate and final outcomes of the project.