PESCaDO

FP7-248594

Personalized Environmental Service Configuration and Delivery Orchestration

D8.2 Roadmap for the development of the PESCaDO workbench

Due date of deliverable: 30.06.2010
Actual submission date: 05.07.2010

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

Lead contractor for this deliverable: FMI

D8.2	Roadmap for the development of the PESCaDO workbench
Project Acronym :	PESCaDO
Contract No :	FP7 - 248594
Due Date :	30.06.2010
Reply To:	Ari Karppinen ari.karppinen@fmi.fi
Actual date of delivery:	05.07.2010

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	D8.2
Deliverable name	Roadmap for the development of the PESCaDO workbench
Type	Report
Status & version	Submitted, V1
Number of pages	46
WP / Task responsible	FMI
Other contributors	UPF, FBK, IOSB, USTUTT, CERTH
Author(s)	All partners of the Consortium
Internal Reviewer	All partners of the Consortium were involved in the cross reviewing of individual sections.
EC Project Officer	Manuel Monteiro
Abstract	The document describes the timeline and resources needed for the development of each functional module of the PESCaDO workbench, and outlines the process of building up the demonstrator system in more detailed and practical terms then in the original work plan.

Table of Contents

Executive Summary

The document lays down the roadmap for PESCaDO workbench development illustrating the progressively increasing functionality of the individual modules as well as of the platform as a whole. It also details the temporal and functional synchronization of the individual modules to ensure this functionality and the timeline and resources that will be needed to achieve this functionality. Together with the original, more general workplan the roadmap will assist in following and controlling the project progress during the development stages of the demonstrator and together especially with deliverable D1.1. will greatly assist in observing and reacting to any problems or deviations from original plan in the final demonstrator development process.

1 Introduction

The purpose of this document is to outline the roadmap for the development of the PESCaDO platform, i.e., to align both short term objectives and long term objectives as specified in the Work Plan with the activities within the individual WPs on the technologies that meet these objectives and their timelines. This implies the specification of:

(i) the approaches to the investigation and realization of the individual modules of the platform and their components as addressed in the individual WPs;
(ii) the progressively increasing functionality (in particular at the time points of the milestones) of the individual modules, their components and the platform as a whole;
(iii) the temporal and functional synchronization of the activities on the modules and the platform to ensure the required functionality;
(iv) the resources that will be needed to achieve this functionality.

The outline of the approaches (without that the content of D8.3 is anticipated) is appropriate in this context since only then can the effort for achieving the required functionality be properly assessed; furthermore, in several WPs alternative or complementary models are being developed such that an adequate distribution of the resources is only possible when the activities to realized these models are outlined.

The outline of the approaches will be combined with the presentation of the activities that will lead to an increasing functionality of the individual modules, broken down to components along the timeline.

An important aspect of this deliverable is thus to outline the timeline and resources needed for building up each functional module, and describe the whole process of building up the demonstrations system in more detailed and practical manner then in the original Work Plan.

The document focuses in describing the resources needed for building each required functional component of the final demonstrator; although the work of building each module is in most cases tightly connected to the corresponding work package and task definitions, investigating the system build-up process from the practical and functional perspective gives a better insight on the practical work and resources needed to achieve the ambitious goals of the project.

As a result, the deliverable will help to follow the project progress during the development stages of the demonstrator and together especially with deliverable D1.1 greatly assist in observing and reacting to any problems or deviations from original plan in the final demonstrator development process.

2 Overview of the Objectives and the Architecture of the Platform

In order to prepare the grounds for the functional description of the individual modules and their components, let us summarize the scientific and technological objectives of PESCaDO as outlined in the Work Plan and display the preliminary working architecture of the system.
The scientific objectives are as follows (cf. also pp. 5-7 of the Work Plan for a more detailed presentation):

SO1. To develop techniques for **discovery of environmental service nodes** in the web.
SO2. To develop **dynamic environmental service uncertainty metric derivation techniques**.
SO3. To define **a protocol infrastructure for orchestration of environmental services**.
SO4. To develop **techniques for orchestration of environmental services**.
SO5. To research and develop technologies for **environmental ontology alignment and extension**.
SO6. To research and develop **techniques for content distillation from multilingual material**.
SO7. To research and develop **user-oriented decision support strategies**.
SO8. To research and develop **visual user-interaction technologies** in the context of user-oriented environmental services.
SO9. To research and develop technologies for **multilingual and multimodal user-tailored environmental information delivery**.

The technological objectives are as follows (cf. also pp. 7-8 of the Work Plan):

TO1. To build up **an operational infrastructure**, **an environmental background knowledge base, a database with auxiliary information** and **stable data delivery channels**.
TO2. To carry out **empirical studies on the needs of users** with respect to environmental information coming from distributed environmental service nodes and **implement a user profile typology** which can be maintained via a flexible user modeling environment.
TO3. To **validate the technologies** developed in PESCaDO by **implementing a number of representative scenarios** of **pilot use cases for the involvement of different types of users in environmental service configuration**.
TO4. To **deliver an operational demonstrator**.

The working architecture of the PESCaDO platform that incorporates the modules whose research and development is targeted by the individual scientific objectives is as sketched in Figure 1. While details of this architecture may change in the course of the project, it serves well as orientation during the first phase of the project.

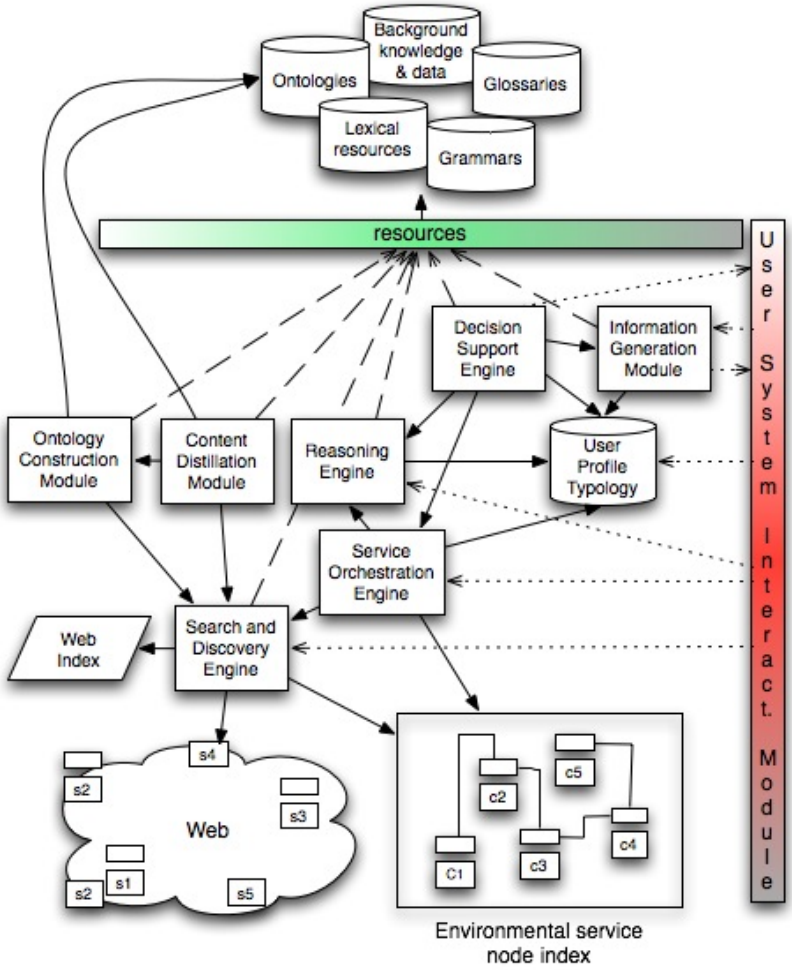


Figure 1: Preliminary working architecture

The following eight main modules are distinguished:

(i) the ontology maintenance and acquisition, i.e., alignment and extension, module (OM);
(ii) the content distillation module (CDM),
(iii) the search and discovery engine (SDE),
(iv) the service orchestration engine (SOE),
(v) the reasoning engine (RE),
(vi) the decision support engine (DSM),
(vii) the information/support generation module (GM), and
(viii) the user-system interaction module (USIM).

This list highlights the close relationship between the scientific objectives and the work on the PESCaDO platform.

The modules will use one or several of the following resources:

(a) the web index,
(b) environmental service node index compiled by the DE
(c) ontologies + background knowledge and data + multilingual term glossaries + lexica and grammars.

(d) a further central resource is the user profile typology, which will be used by the decision support, information/support generation and reasoning modules, and which will also be accessible by the user.

The central module to interact with the user will be the DSM; the user will be able to solicit support or information invoking this module, using for the formulation of their request the graphical problem description language interface. With the user profile at hand (accessed in the user profile typology), DSM will initiate the RE and, if necessary, the SOE. Operating on the service node index and calling the SDE, the SOE orchestrates the optimal node configuration to serve the user's request, using the uncertainty metrics to assess the performance of the individual nodes and of the configuration as a whole.

3 Functional Description of the Modules

In this section, the modules of the PESCaDO platform and their components are outlined from the functional perspective. The organization of the section follows the structure of the Work Plan in terms of Work Packages.

3.1 Environmental service node discovery

The development of the environmental node service discovery module will consist of the following activities; each of them corresponding to one or two subtasks in the Work Plan of WP 2:

(i) set up of the search infrastructure (T2.1);
(ii) empirical study of the web pages of a collection of web-based environmental service nodes (T2.2);
(iii) development of the query formulation and expansion component (T2.3);
(iv) development of the repository initialization and node indexing components (T2.4 and T2.6);
(v) development of the environmental node search module (T2.5).

3.1.1 Environmental node search infrastructure setup

The search infrastructure setup will deal with the architecture design, the development and the integration of several components that will enable domain specific search. The following components will be integrated in this architecture:

(i) web search API: a web search application, which will be employed to enable web access;
(ii) web crawler: a computer application that browses the World Wide Web in a methodical, automated manner; web crawlers will be used to expand the results of the web search engines;
(iii) classifier: a component that it is trained with ground truth information and is capable of classifying new data into different classes; in this case a binary classifier will be considered to distinguish relevant from non-relevant information;
(iv) query expansion module: a module that enables query expansion; cf. Subsection 3.1.3 below for a description;

(v) search module: a module that enables search into the web service nodes; cf. Subsection 3.1.4 below for a description;
(vi) content distillery module: this module is responsible for extracting information from the service nodes; its development is described in WP4 (cf. Subsection 3.3 below).
(vii) node repository: this component will serve as a storage structure, in which the node access information will be maintained; the repository is to be described in Subsection 3.1.5.

Dependencies

The external WP dependencies concern the content distillery module and the data retrieval service that will directly access the repository. Especially the content distillery module plays a very important role as it is considered as an integral part of the architecture. The design of the architecture should be also based on the current state of the art regarding domain-specific search as reflected in WP8.

3.1.2 Empirical study

During this activity, an empirical study of a representative collection of web pages and web services offered by environmental services will be carried out. The goal of the study is to assess the specifics of the metadata and content codification in web pages containing environmental material in order to be able to design efficient indexing and content distillery strategies.

Dependencies

The completion of this task strongly depends on the availability and the quality of the collection of the web pages that will be provided by the environmental experts in the Consortium, i.e., FMI and HSY.

3.1.3 Query formulation and expansion module implementation

This activity deals with the design and development of multilingual query formulation and expansion techniques for the search of environmental service nodes for a specific region. More specifically the following subcomponents can be identified:

(i) Keyword spice extraction and query expansion subcomponent
(ii) Inference and reasoning based query expansion subcomponent
(iii) Interactive query formulation subcomponent

Dependencies

The development of the query formulation and expansion techniques and the corresponding module require the results of the empirical study. In addition, in order to enable query expansion based on inference and reasoning strategies we depend on the environmental ontologies that are created in WP 4 (cf. Section 3.3). To support interactive interfaces for the query formulation and expansion, there is a strong dependency with the user interfaces developed in WP 6 (cf. Section 3.5).

3.1.4 Search Module Development

This task is dedicated to the design and development of a search module which will be capable of searching the environmental nodes. It consists of the following submodules:

(i) Keyword-based text and metadata search submodule
(ii) Semantic search submodule
(iii) Web page language identification submodule

Dependencies

This task depends on the ontologies created in WP4 in order to enable the semantic search, as well as on the text extraction techniques of WP4 that will facilitate keyword based search and language identification.

3.1.5 Repository initialization and node indexing

This task deals with the design and development of techniques for functional indexing of the retrieved environmental nodes with a view to setting up a functional index repository in which the discovered environmental nodes are indexed with respect to their metadata-based characterization.

Dependencies

The repository would be a storage structure, in which the results extracted with the aid of the aforementioned modules will be stored. The major dependency that can be reported has to do with the empirical study as the fingerprint construction will mainly be based on these findings.

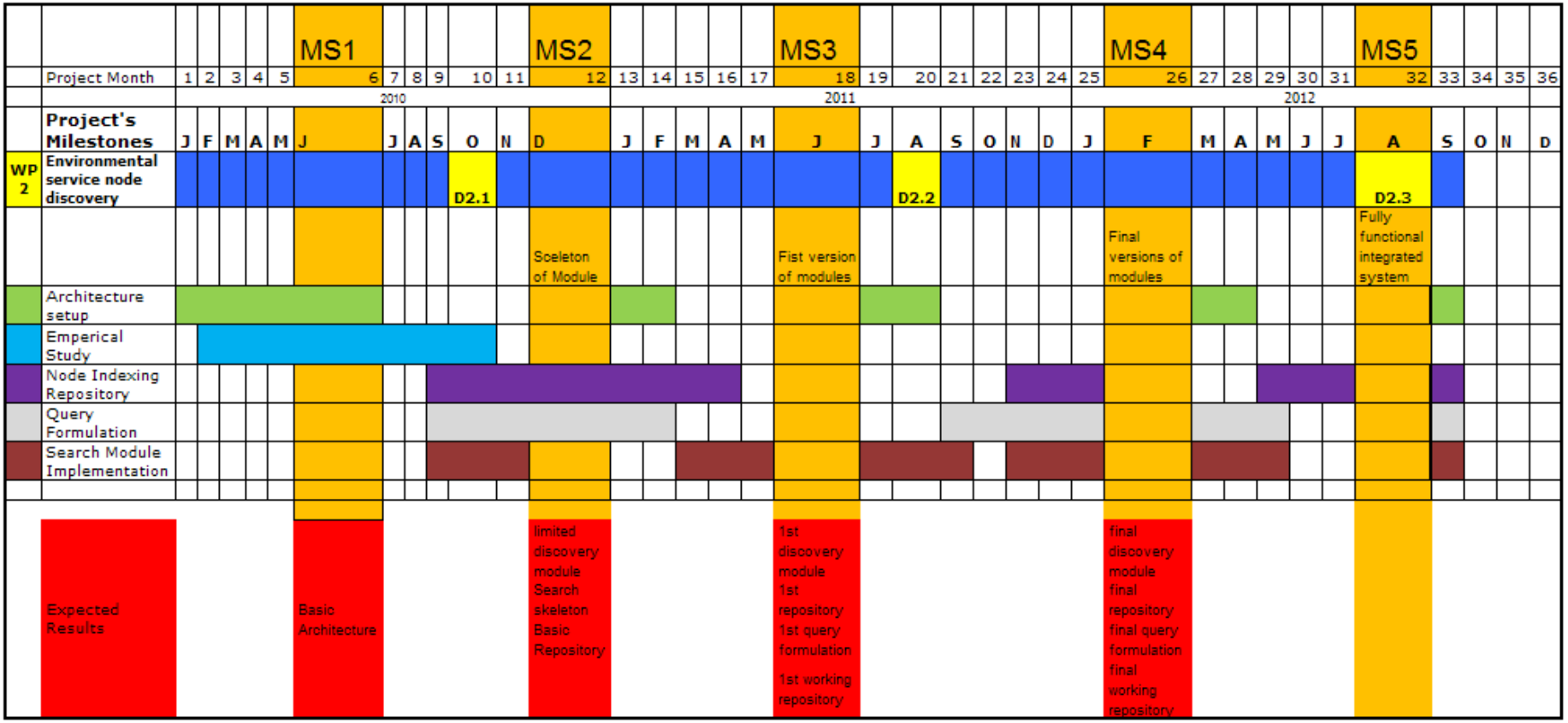
3.1.6 Activity Table and Workload

In the following table, we specify the workload distribution in PMs across the tasks and subtasks and summarize the external and internal resources needed to carry out them in order to develop the corresponding modules and their components.

Activity	Subactivity	Description	Dependencies	Workload
Search Infrastructure Setup	Web Search API development	A web search application, which will be employed to enable web access		2
	Web Crawler componpent development	A computer application that browses the World Wide Web in a methodical, automated manner.	Results of empirical study Training samples	3
	Classifier component development	A component that it is trained with ground truth information and is capable of classifying new data into different classes.	Training samples. Results of empirical study Content Distillery techniques of WP4	3
Empirical		The goal of the study is	Availability of a	5

Study		to assess the specifics of the metadata and content codification in web pages containing environmental material.	quality representative sample of environmental nodes	
Query formulation and expansion	Keyword spice extraction and query expansion module development	This module will be capable of expanding the initial query by generating domain specific keywords	Empirical study Content Distillery techniques of WP4	4
	Inference and reasoning based query expansion module development	This component will access the ontologies and expand the query by performing reasoning and inference	Ontologies of WP4	3
	Interactive query formulation module development	This module will consist of an interactive interface that will support manual assisted query formulation and expansion	User interactive interfaces of WP7 Development of use cases for query formulation/expansion	4
Environmental node search module development	Keyword-based text and metadata search module	This component would support keyword search into the environmental nodes	Content Distillery techniques of WP4	3
	Semantic search module development	This module would support concept based search in the environmental nodes	Ontologies of WP4	3
	Web page language identification component development	This component will be capable of recognizing the language of the retrieved environmental node.	Content Distillery techniques of WP4	4
Repository initialization and Node indexing components development		Development of a functional index repository in which the environmental nodes are indexed with respect to their metadata-based characterization	Results of empirical study	9

3.1.7 Timeline



3.1.8 Critical Factors

The critical factors for the development of the Environmental Node Service Discovery Module are:

1. The availability of corpora both for the empirical study and for training of machine-learning based algorithms (classifiers, crawlers, keyword spice, etc.).
2. The availability of the ontologies of WP4 to perform reasoning- and inference-based actions.
3. Integration of content distillery techniques developed in WP 4 in the WP2 architecture.
4. Quality of results in the case of extracting from an environmental node and matching it with the information extracted from other nodes.
5. A reasonable compromise between the time response of the system and the cardinality of the retrieved results should be performed in order to achieve an optimal recall and precision.

3.2 Orchestration of Environmental Services

The activities related to the development of the module for the orchestration of environmental services consist in:

(i) the design and realization of the infrastructure for the interconnection of environmental services (T3.1);
(ii) development of uncertainty metrics for the assessment of the quality of the output of individual environmental service nodes (T3.2);
(iii) development of matchmaking strategies (based on the uncertainty metrics) between the environmental service nodes researched by the search engine developed in WP 2, with the goal to find the nodes that provide the most adequate and most reliable information needed to handle the requests of the decision support and user interaction techniques (T3.3);
(iv) development of a component for connecting the selected nodes to establish the data, knowledge and control flow (T3.4).

3.2.1 Infrastructure for the interconnection of environmental services

The activity on the realization of the infrastructure for the interconnection of environmental services will address the following tasks:

(i) review of the state-of-the-art data- and semantics-oriented protocols used for interconnection of environmental services as well as analysis (in coordination with activities in WP2) of the access to the information of environmental nodes collected and studied in WP2 (cf. 3.1.2 above), i.e., data formats of webpage-based services and web service interfaces of nodes that are accessible via web services;
(ii) design of the interconnection strategies of environmental service nodes drawing upon the results of the first task;
(iii) implementation of web services and, where appropriate, local interfaces, for the interconnection of environmental service nodes.

Dependencies

This activity depends, first of all, on WP2 (activity 3.2.1 above) and WP4, as far as potential references to ontologies within the interfaces are concerned.

3.2.2 Uncertainty Metrics

The activities related to the development of uncertainty metrics address the problem from two angles: (a) imprecision metrics that assess the uncertainty of the output of services stemming from the lack of erroneous data and data gaps, and confidence metrics that assess the uncertainty of the services' output from the lack of trustworthiness of the services themselves. These activities comprise:

(i) study of the state-of-the art imprecision metrics and exchange of ideas with the UncertWeb project (FP7, Objective 6.4b), which focuses on imprecision metrics;
(ii) empirical study of the quality and granularity of the output of a representative set of environmental nodes in the web with the goal to provide the basis for the determination of correlations between the characteristics of these nodes and the quality of their output;

(iii) implementation of the automatic derivation of parameter sets for the classification of environmental nodes with respect to their trustworthiness and derivation of confidence metrics;
(iv) implementation of the imprecision metrics.

Dependencies

This activity depends on an early availability of a representative set of web-based environmental services and on the successful collaboration with UncertWeb.

3.2.3 Matchmaking strategies

In the scope of this task, standard service matching strategies for opportunistic collaboration known from communication technologies will be examined and the most suitable one will be adapted for matchmaking of environmental service nodes. The integration of the uncertainty metrics into the matchmaking procedure will be essential.

Dependencies

This activity depends on the prior availability of the uncertainty metrics and the results of the study of the matchmaking strategies.

3.2.4 Development and implementation of techniques for ontology based connecting services

The activities related to the development of the connecting services are planned to directly incorporate the results of the matchmaking strategies. Their main objective will be to level out the output of services to be used either as input to another service (in the case of chaining) or as complementary / competing information with respect to the output of another service in order to make it compatible with the required input or output, respectively of other services. This will be achieved by using ontologies as intermediary representation.

Dependencies

This activity depends on the prior availability of the matchmaking strategies and, externally, on the availability of ontologies (cf. WP 4; Subsection 3.3 below).

In what follows, we specify the workload distribution across the development of the individual submodules and components (in PMs) and summarize the external and internal resources needed by these submodules / components.

3.2.5 Activity Table and Workload

In what follows, we specify the workload distribution across the development of the individual submodules and components (in PMs) and summarize the external and internal resources needed by these submodules / components.

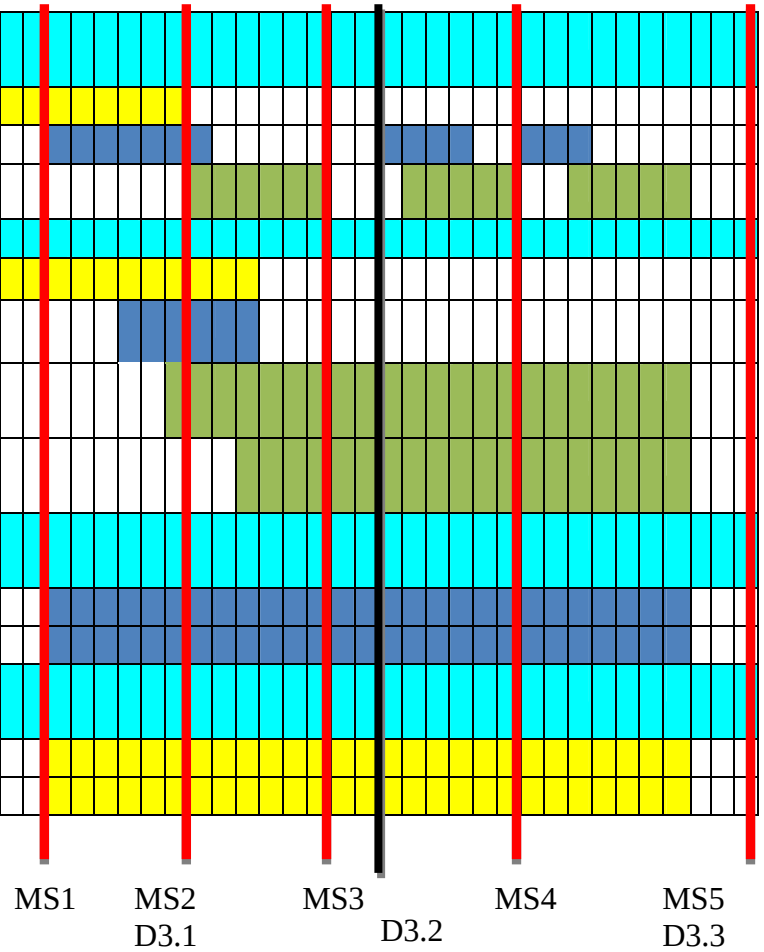
Activity	Subactivity	Description	Dependencies	WL
Infrastructure for the interconnection	Review of the protocols; analysis	Review of data- and semantics-oriented protocols used for interconnection of environmental	WP2 WP4	2

of environmental Services		services and the access to the information of environmental nodes		
	Design of interconnection strategies	Design of the interconnection strategies of environmental services	review & analysis	2
	Implementation of web services and local interfaces	Implementation	Previous tasks	3
Uncertainty metrics	Review of the state-of-art ; co-op with UncertWeb	Study of the state-of-the art imprecision metrics and exchange of ideas with the UncertWeb project,	Availability of services; UncertWeb	4
	Empirical study	Study of the quality and granularity of a representative set of environmental nodes in the web providing the basis for the determination of correlations between the characteristics of these nodes and the quality of their output		4
	Implementation of the automatic classification of the nodes	Implementation of the automatic derivation of parameter sets for the classification of environmental nodes with respect to their trustworthiness and derivation of confidence metrics;	.	6
	Implementation of the impercision metrics	Implementation		6
Matchmaking	Study the existing matchmaking strategies	Standard service matching strategies for opportunistic collaboration known from communication technologies will be examined	Availability of uncertainty metrics	2
	Selection and implementation of the strategies	Implementation	Uncertainty metrics; study results	5
Development and implementation of techniques for ontology based connecting services	Techniques for information chaining	Level out the output of services to be used as input to another service in order to make it compatible with the required input or output of other services.	Availability of matchmaking strategies. WP4	3
	Techniques for information merge and fusion	level out the output of services to be used as complementary/ competing information with respect to other services	Availability of matchmaking strategies. WP4	5

3.2.6 Timeline and dependency from other modules

The following table shows the timeline for the development of the individual submodules and components and their temporal dependencies on work carried out on other modules (mainly in other WPs).

Tasks	Interdependency	1st year 2010	2nd year 2011	3rd year 2012
Infrastructure for the interconnection of environmental services				
Review of the protocols; analysis	WP2 & WP4			
Design of interconnection strategies	review & analyses			
Implementation of web services and local interfaces	strategy design			
Uncertainty metrics				
Review of the state-of-art ; co-op with UncertWeb				
Empirical study				
Implementation of the automatic derivation of parameter sets for classification of nodes	review,co-op , empirical study			
Implementation of the impercision metrics	all earlier uncertainty metrics submodules			
Matchmaking between the environmental service nodes				
Study the existing matchmaking strategies				
Selection and implementation of the strategies	Service node discovery (SND)			
Development and implementation of techniques for ontology based connecting services				
Service chaining	ontology module (WP4)			
Information merge and fusion	ontology module (WP4)			



3.2.7 Critical factors

The success in successfully identifying and implementing techniques for transferring the various environmental information available in the web to a form which allows fusion, merge, chaining and uncertainty analyses of the services will be the critical perquisite of the service orchestration.

After that, the definition of a generally applicable metrics which is quantitatively able to classify the reliability and uncertainty of wide variety of environmental services will be the main challenge. It is essential to find a sufficient set of "base" services, whose uncertainty can be properly and quantitatively assessed and which are representative enough to provide information for majority of the user cases.

3.3 Environmental ontology construction and population

Aim of the module is to guarantee the availability of ontology modules that ensure that

(i) during the composition of distributed environmental node configurations the input/output of the individual nodes can be "leveled out", with the ontologies serving as an "interlingua" on which representations of different abstraction can be projected, respectively derived;
(ii) the pilot use case domain ontologies are sufficiently comprehensive in order to allow, on the one hand, for reasoning on the ontologies with the goal to derive adequate content for user decision support, and, on the other hand, for generation of user-tailored environmental information.

The environmental ontology population and building module will consist of two major submodules: (i) the environmental ontology construction and (ii) the content distillation submodule. Each of these submodules will contain, in turn, several components.
The following activities address their development:

(i) empirical study of the available environmental ontologies and collection/assessment of available environmental corpora from which PESCaDO can distil relevant information for the use cases (T4.1 and T4.2);
(ii) setting up of the linguistic corpus processing infrastructure for information distillation (T4.3);
(iii) development of techniques for automatic alignment and extension of environmental ontologies (T4.4);
(iv) development of techniques for content distillation from multilingual environmental material.

3.3.1 Environmental ontology construction

This sub-module has the aim of building the ontological resources that are needed for the aims of PESCaDO. Such resources will be built resorting to a mixture of approaches and techniques: 1. re-use of existing ontologies; 2. automatic reasoning techniques to extend and integrate them; 3. content distillation techniques to automatically extend existing ontologies and to check their coverage; 4. manual construction of ontologies.

Dependencies

The complexity of this activity depends on the completeness of the available environmental ontologies. Furthermore, it depends on WP5 for reasoning-based extension of ontologies.

3.3.2 Content distillation module

The content distillation sub-module will develop techniques for the extraction of content from texts in the environmental domain for English, Swedish, and Finnish. The extraction techniques will range from shallow key-phrase extraction techniques to deeper frame-semantics based techniques. The extracted content will be used as a support to environmental node discovery (WP2), for ontology learning, and for populating the knowledge base with all the environmental data which are needed for the PESCaDO service.

Dependencies

This activity relies to a major extent on FBK's linguistic workbench that is made available to the project.

3.3.3 Activity Table and Workload

Activity	Subactivity	Description	Dependencies	Workload
Environmental contology construction module	Empirical study of available environmental ontologies	Determine the material that is already available and what should be used , extended and merged in PESCaDO		1
	Collection of multilingual text corpora in the environmental domain	Collect available documents and HTML pages on the environmental domain, for English, Swedish and Finnish (will be used also in WP7).	Availability of tools for the automatic collection of domain-specific documents from the Web. Availability of a sufficient number of documents/web pages on the Web.	2
	Ontology merging and alignment	Merging of the collected ontologies to check missing or overlapping parts	Availability of existing ontologies	3
	First ontology extension cycle	Using content distillation techniques for the creation of new concepts and their integration in the ontologies	Requires the development of a set of content distillation techniques and their previous evaluation (see sub-module 1.2)	4

	Manual correction and validation of acquired ontology concepts	A general check of the outcome of the previous sub-activity is carried out as well as the manual creation of parts of the ontology which were not automatically created	Previous sub-activity should be complete	2
	Ontology integration and corpus based coverage check	Reasoning techniques (WP5) are used to automatically find related concepts from ontologies of different abstraction; corpus-based linguistic analyses are used to check the coverage of the enriched ontologies	Availability of a reasoning module. Availability of a corpus for statistical and linguistic analysis (see sub-activity 2) and of natural language processing tools.	4
	Manual correction and validation of acquired ontology concepts	A general check of the outcome of the previous sub-activity is carried out as well as the manual creation of missing parts of the ontology which were not automatically created	Previous sub-activity should be complete	3
Content distillation module	Development of keyphrase-based knowledge extraction tools	Develop a system that can extract key-concepts from environmental texts in all project languages. The output of this sub-activity will be used by the ontology construction sub-module and the node discovery module (WP2)	Availability of a domain-specific multilingual corpus (subactivity 4.1.2) in order to tune the parameters required for keyphrase extraction	3
	Development and testing of different techniques for relation extraction	Identify the most relevant relations to be completed or integrated in the environmental	Availability of a domain-specific multilingual corpus (subactivity 4.1.2) in order to extract the relations.	3

		ontologies and develop methodologies for their automatic extraction from corpora, the Web, WikiPedia and/or WordNet		
	Creation of new ontology concepts and their integration in the existing ontologies	Use the relation extraction techniques and the terms from domain corpora (sub-activity 4.2.1 and 4.2.2.) to create new concepts and integrate them in the existing ontologies	Completion of 4.2.1. and 4.2.2. subtasks, i.e. availability of keyphrase extraction and relation extraction applications	2
	Development of frame-based approaches for deep-content distillation from multilingual material	Test rule-based and statistical techniques for deep-content distillation based on semantic role labelling	Availability of a domain-specific multilingual corpus (subactivity 4.1.2)	4
	Apply deep content distillation approaches to advanced ontology extension	Apply the algorithms developed in the previous sub-activity for advanced ontology extension	Use of the existing ontologies collected and merged in the previous module	2
	Apply deep content distillation approaches to the extraction of information from web-pages of environmental nodes	Apply the algorithms developed in the previous sub-activity for extracting relevant information from environmental nodes	To be used in WP2, T2.5.2.	2

3.3.4 Timeline

The two modules (4.1 and 4.2) will be available in a preliminary version for the first prototype (month 18) applying basic techniques, while a more complex and complete version will be integrated in the second prototype (month 24). However, the knowledge acquisition procedures will go on and will be improved until month 34.

Tasks	Interdependencies	1st year 2010	2nd year 2011	3rd year 2012
4.1 Environmental ontology construction				
4.1.1. Empirical study of available environmental ontologies				
4.1.2. Collection of multilingual text corpora				
4.1.3. Ontology merging and alignment				
4.1.4. First (basic) ontology extension cycle				
4.1.5. Manual correction and validation of acquired ontology concepts	Completion of 4.1.4.			
4.1.6. Ontology integration and corpus based coverage check	Use reasoning techniques developed in WP 5 and deep-content distillation techniques (4.2.4)			
4.1.7. Manual correction and validation of acquired ontology concepts	Completion of 4.1.6.			
4.2 Content distillation module				
4.2.1. Development of keyword-based knowledge extraction tools	Availability of domain corpora (4.1.2.)			
4.2.2. Development and testing of different techniques for relation extraction	Availability of domain corpora (4.1.2.)			
4.2.3. Creation of new ontology concepts and their	Availability of			

integration in the existing ontologies	domain ontologies to enrich (4.1.3.). Availability of ontology extension tools (4.2.1)																																		
4.2.4. Development of frame-based approaches for deep-content distillation from multilingual material	Availability of multilingual corpora (4.1.2.).																																		
4.2.5. Apply deep content distillation approaches to advanced ontology extension																																			
4.2.6. Apply deep content distillation approaches to the extraction of information from web-pages of environmental nodes	To be used in WP2, T2.5.2.																																		

3.3.5 Critical Factors

A critical factor which may affect the construction of environmental ontologies is the availability of domain-specific texts to be used in the relation extraction step. In particular, since the information needed is very specific (the relationship between environmental issues and health conditions), it may be difficult to collect large corpora dealing with this specific topic.

Content distillation techniques should deal with English, Finnish and Swedish with a comparable accuracy. While a wide range of tools and resources are available for English, the two other languages may require complex adaptations. For example, the training of statistical natural language processing systems requires large annotated data, which are not available for Finnish and Swedish and cannot created within the project. As a fall-back strategy, we will experiment with cross-language information extraction techniques, based on available automatic translation tools.

3.4 Reasoning and user tailored decision support

The activities related to the development of the reasoning and user tailored decision support consist in:

(i) working out of the problem description language (PDL) in which the user can specify his request for decision support and submit specific questions concerning environmental conditions (T5.1);
(ii) develop user-assisted reasoning techniques to deduce the knowledge required to meet the user's requests (T5.2);
(iii) develop user-oriented and problem-driven content selection techniques to facilitate the adequate information to the user (T5.3).

3.4.1 Problem Description Language Specification

The R&D activities associated with the specification of the problem description language (PDL) will consist in:

(i) the analysis of the state of the art of PDL specifications;
(ii) the analysis of the needs of the user in the PESCaDO applications with respect to decision support, the specification of an OWL-axiom based PDL; and
(iii) the development of a system module, namely the Problem Description Generation Module, responsible of assembling, from the request of the user (acquired via a PDL-user graphical interface, to be developed in coordination with WP6), a valid problem description to be submitted to the system.

Dependencies

The problem description language specification will depend on the availability of a description of (and will be designed to cover) the use cases considered in PESCaDO.

The Problem Description Generation Module will interact with the user interface to get the data required to put together a problem description. The problem description will then be processed by other modules of the system (e.g. the Decision Module) in order to fulfil the user request.

3.4.2 Reasoning Module

This task will build the core semantic reasoning services that are necessary in order to fulfil the requests of the users as well as the requests for reasoning by the other modules. We plan to base the reasoning strategies on a combination of ontological reasoning (on the ontologies made available in WP4) and content filtering,
In addition to the investigation of the state of the art on user-oriented reasoning support, the module will consist of two main activities:

(i) definition and development of reasoning services to determine the knowledge needed to fulfil a user request;
(ii) definition and development of reasoning services to infer new knowledge based on asserted knowledge and the data retrieved from environmental service nodes.

In particular, we plan to investigate and apply a combination of

(a) OWL DL-based reasoning;
(b) rule-based reasoning;
(c) reasoning on probabilistic and uncertain models (based e.g. on Bayesian Networks or Markov Logics).

Interaction with /Dependency from other modules

Being one of the key components of the system, the reasoning module is closely intertwined with several other modules; among them:

- the query formulation and expansion module developed in WP2;
- the environmental service nodes search modules developed in WP2;
- the ontology based connecting services module developed in WP3;
- the content selection module developed in WP5;
- the user interface component developed in WP6.

This module clearly heavily relies on the availability of:

- the environmental and environmental-related ontologies produced by WP4;
- the problem description returned by the Problem Description Generation Module;
- the data collected from the environmental service nodes.

3.4.3 Content Selection Module

This module has the objective to filter the knowledge derived by the Reasoning module according to the user preferences/profile and the concrete decision support request, producing a content plan of the knowledge to be communicated to the user in order to offer a reasonable decision support and satisfy the user's needs. Different content selection strategies will be researched for the different types of the material expected by the user from the system: reports, instructions (suggestions, recommendations and indications), and warnings.
In particular, we plan to focus our R&D work on:

(i) definition and development of reasoning service tailored to content selection activities;
(ii) definition and development of content selection strategies based on learning reinforcement approaches; both on-line and off-line learning reinforcement strategies will be investigated, and the most suitable one will be implemented;

(iii) definition and development of user-interaction facilities that allow the expert/user to influence the content selection strategy.

The R&D work performed within this module will be performed in tight coordination with Task 7.2 in WP7.

Dependencies

This module is closely intertwined with the information production module, which produces the actual text, table, and graphic content to be communicated to the user. The reasoning services for content selection will be implemented taking in consideration the work on the Reasoning module.
The other module, with which the content selection module will interact, is the user interface module: the content selection module will use it to enable experts/end users interaction with the content selection strategies.

3.4.4 Activity Table and Workload

In what follows, we specify the workload distribution across the development of the individual modules and components (in PMs) and summarize the external and internal resources needed by these submodules / components.

Activity	Subactivity	Description	Dependencies	Workload
Problem Description Language Specification	Analysis of the state of the art of PDL specifications	Availability of other OWL-axiom based PDLs will be investigated		1
	Analysis of the needs of the user in the PESCaDO applications with respect to decision support	An analysis of the possible user requests according to the use cases considered in the project will be performed	Definition of the pilot use cases	2
	Specification of the PDL	Actual specification of the PDL (based on OWL)		3
	Development of the problem description generation module and user-graphical interface	Development of a module which acquires from the user his/her request, and generates a valid problem description to be processed by the system	Availability of the graphical user interface from WP6	2
Reasoning Module	Analysis of the state of the art of user oriented reasoning support	We will investigate the availability of state of the art reasoning system, in particular those combining ontological reasoning and statistical/probabilistic reasoning		3
	Definition and development of reasoning services to determine the	We will implement techniques based on OWL DL reasoning. If the situation will require it,	Availability of: ontologies describing the domain, the	8

	knowledge needed to fulfil a user request	rule-based strategies will also be applied.	problem description structure, user profile	
	Definition and development of reasoning services to infer new knowledge	We will implement techniques based on OWL DL reasoning. rule-based reasoning, and probabilistic/statistical reasoning.	Availability of: ontologies describing the domain, the problem description structure, user profile, data collected from environmental service node	17
Content Selection Module	Definition and development of reasoning service for content selection	Investigation and implementation of reasoning techniques for supporting the selection of content to be shown to the user.	Availability of the content inferred by the reasoning module, user profile, problem description, domain ontologies	4
	Definition and development of content selection strategies based on learning reinforcement models	We will investigate both on-line and off-line learning reinforcement strategies for content selection. Content selection strategies for reports, instructions and warnings will be implemented	Availability of the content inferred by the reasoning module, user profile, problem description, domain ontologies	4.5
	Definition and development of user-interaction facilities for content selection	Implementation of a user interface/facilities to support the interaction of expert users with the content selection strategies	Definition of the content selection strategy, availability of the content inferred by the reasoning module, user profile, problem description, domain ontologies	4

3.4.5 Timeline and interdependencies

The following table shows the timeline for the development of the individual submodules and components and their temporal dependencies on work carried out on other modules (mainly in other WPs).

Tasks	Interdependency	1st year 2010	2nd year 2011	3rd year 2012
Problem Description Language Specification				
Analysis of the state of the art of PDL specifications				
Analysis of the needs of the user in the PESCaDO applications with respect to decision support	WP8 (pilot use casess)			
Specification of the PDL				
Development of the problem description generation module and user-graphical interface	WP6 (user interface)			
Reasoning Module				
Analysis of the state of the art of user oriented reasoning support				
Definition and development of reasoning services to determine the knowledge needed to fulfil a user request	WP2 to WP6			
Definition and development of reasoning services to infer new knowledge	WP2 to WP6			
Content Selection Module	WP7 (Task 7.2)			
Definition and development of reasoning service for content selection				
Definition and development of content selection strategies based on learning reinforcement models				
Definition and development of user-interaction facilities for content selection	WP6 (user interface)			

MS1 MS2 MS3 MS4 MS5

3.4.6 Critical Factors

The critical factors for the Problem Description Language Specification are:

- the availability of the description of the pilot use cases, to derive a general enough formulation of the user requests;

The critical factors for the Reasoning Module are:

- the variety of data sources on which to work on, and their availability: environmental domain ontologies, relations connecting the various environmental domain ontologies, data obtained from the environmental service nodes, user profiles, and problem descriptions.
- the size of the knowledge base on which to reason upon;
- the challenge of applying relatively recent work combing probabilistic/statistical reasoning with logic-based reasoning in a real world application;

The critical factors for the Content Selection Module are:

- the availability of the inferred knowledge produced by the Reasoning Module;
- the size of the knowledge base, which may affect the decision of implementing on-line learning reinforcement strategies rather than off-line ones.

3.5 Visualization and user interaction

The activities related to visual interaction are concerned with the development of components that form part of a number of different modules across the PESCaDO platform, namely:

(i) uncertainty metrics derivation, using visual analytics methodology (T6.1)
(ii) environmental service node discovery (T6.2);
(iii) environmental service orchestration (T6.3);
(iv) user-oriented and problem-driven decision support (T6.4).

Visual interaction activities are thus transversal to all other R&D activities in PESCaDO.

3.5.1 Visual analytics techniques for uncertainty metrics determination

This activity targets the development of a visual analytics component for uncertainty metrics derivation. It consists in:

(i) study of state-of-the-art visual analytics methodologies;
(ii) selection and adaption of the most suitable methodology to the context of PESCaDO in order to facilitate the display and interactive modification of the most important parameters of the metrics derivation algorithms.

Based on early design decision in the orchestration module (cf. Subsection 3.2 above) we will select and adapt appropriate interactive techniques to show and explore the important aspects of the results of the metrics derivation algorithms, to refine their parameters, and to close the feedback loop by re-evaluating the derivation algorithms.

User interaction will involve the intervention by qualitative feedback of the result of the application of the metrics, by assigning confidence marks (from a predefined qualitative scale) to specific environmental nodes in the web. First prototypes will be built with Java and the information visualization toolkit *prefuse*.

The resulting VA techniques will be iteratively evaluated and redesigned in close collaboration with FMI. Subsequent activity for visual analytics support for metric determination will be situated in Task 6.3.

Dependencies

There is a direct dependency with the activity related to Task 3.2 (WP 3).

3.5.2 Interactive techniques for the process of environmental node discovery

The formulation of queries to discover new environmental nodes in the web is not a research task within the project but an administrator task within the PESCaDO system. Hence, user intervention will be supported during the procedure of query expansion such that the user will be able to modify and extend the queries composed by the system via a graphical interface. This approach shall be based on and supported by relevance feedback mechanisms.

The activity thus consists in:

(i) selection and adaption of machine learning-based visualization techniques such as landscape visualizations for Self Organizing Maps to enable the user to explore results of query, derive insights of the exploration;

(ii) realization of the interactive feedback mechanisms.

Dependencies

There is a direct dependency with the activities of WP2 related to query formulation and expansion (cf. Subsection 3.1.3 above).

3.5.3 Interactive techniques for the process of service orchestration

This activity starts later in the project, because the realization of the interactive intervention of the user in the procedure of node orchestration requires intermediate results from WPs 2 and 3 (mainly Tasks 2.2 and 3.3). Furthermore it is a continuation of the activities related to Tasks 6.1 and 6.2 (cf. Subsections 3.5.1 and 3.5.2 above).

The intervention concerns modification of environmental node exclusion or inclusion criteria and adjustment of the service configurations created by the system. Its realization consists in the following subactivities:

(i) analysis of state-of-the-art node-link based visualization techniques for data provenance;

(ii) adaption of the selected node-link based visualization techniques to the specifics of the environmental service node orchestration such that the user will be able to monitor, and – if necessary or desired – intervene into the process.

(iii) a subsequent evaluation study with test users will be used to assess the suitability of the mechanisms and to determine targeted ameliorations or modifications.

Dependencies

There is a direct dependency with the activities of WP2 and WP3 as well as with the activities described in Subsections 3.5.1 and 3.5.2.

3.5.4 Visual interaction techniques for decision support scenarios

The activity consists in:

(i) Based on the first results from the user requirement analysis (D8.4 due in M8) and the problem description language (PDL) for query representation (D5.1 due in M8) existing visual query editors (e.g. from PATExpert's visualization module) will be adapted to the PDL or new ones will be developed if necessary. End users of the demonstration use case 1 might not benefit of complex query generation capabilities, so that an early, basic prototype implementation based on wizards and forms will be used for the integration effort during system development.

(ii) In collaboration with the information production module of WP7, a visual relevance feedback mechanism will be realized for extended or subsequent use cases involving administrators, power users, or service providers. For this task, an integrated user interface for both, the visual query editor as well as the result representation will be developed, based on the techniques USTUTT developed in the FP6 project PATExpert for graphical search query formulation and feedback provision.

(iii) In order to realize user intervention with the underlying reasoning and result representation modules, additional visualization methods for displaying reasoning trails and machine learning results should be developed and integrated. Here, also a close collaboration with WP4 will be established.

Dependencies

This activity needs to be coordinated with WP 4, 6, 7 and 8.

3.5.5 Activity Table and Workload

We specify here the workload distribution across the development of the individual submodules and components (in PMs) and summarize the external and internal resources needed by these submodules / components.

Activity	Subactivity	Dependencies	Workload
Task 6.1: Development of visual analytics techniques for uncertainty metrics determination	Research the state of the art.	Intertwined with the development of uncertainty metrics derivation algorithms in Task 3.2 of WP3	1
	Design, analyse, and implement VA supported strategy for metrics determination		5
	Evaluation and adaption of the implemented strategies	All previous subactivities of this tasks	2
Task 6.2: Development of interactive techniques for the process of environmental node discovery	Research the state of the art and analyse the needs of the users with respect to query expansion in PESCaDO and with respect to feedback provision.	Empirical Study of WP2	1
	Develop visual presentations for		2

	node discovery results		
	Design interaction methods for query expansion and relevance feedback for node discovery	Query formulation and expansion interfaces of WP2	2
	Evaluation and adaption of the implemented strategies	All previous subactivities of this tasks	1
Task 6.3: Development of interactive techniques for the process of service Orchestration	Develop Visual representations for node orchestration setup	Service infrastructure description of WP3	2
	Design and implement interaction methods for user intervention during node orchestration	Intervention interfaces for matchmaking strategies from WP3	2
	Evaluation and adaption of the implemented strategies	All previous subactivities of this tasks	2
Task 6.4: Development of visual interaction techniques for decision support scenarios	Research the state of the art and analyse user needs for personalized environmental decision support.	none	2
	Design the interactive graphical surface of the PDL	Specification of the user-oriented problem description language (PDL) of WP5	3
	Develop result representations and visual relevance feedback with respect to the quality of the provided decision support	Output structure specification of Information Production module of WP7	3
	Evaluation and adaption of the implemented strategies	All previous subactivities of this tasks	6

3.5.6 Timeline and interdependencies

The following table shows the timeline for the development of the individual submodules and components

Tasks	1st year 2010												2nd year 2011												3rd year 2012											
Month	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
6.1 Visual analytics for uncertainty metrics																																				
6.2 Interactive techniques for env. node discovery																																				
6.3 Interactive techniques for service orchestration																																				
6.4 Interactive techniques for decision support																																				
Deliverables								8.4		8.5 2.1				6.1						6.2														6.3		

3.5.7 Critical Factors

The critical factors for the development of the visualization and user interaction techniques module are the involvement of the user groups and the early establishment of interfaces to other work package results.

3.6 Information Production

The information production module will consist of two major submodules: (i) the discourse planning submodule, and (i) information generation submodule. Each of these submodules will contain, in its turn, of several components. The following activities are foreseen in the context of the development of these submodules:

(i) empirical study of the environmental textual material for the determination of the primary linguistic constructions that need to be captured by the information production module (T7.1);
(ii) design and development of discourse-driven content selection techniques (T7.2);
(iii) design and development of discourse planning techniques (T7.3);
(iv) design and development of mode selection techniques (T7.4);
(v) development of multilingual resources for text generation (T7.5);
(vi) adaption of off-the-shelf graphic and table generators and their integration into PESCaDO's information production module (T7.5).

3.6.1 Discourse planning

The main components of the discourse planning submodule are: (i.1) mode selection component; (i.2) discourse structure planning component.
The mode selection component will be able to select the appropriate presentation mode (text, table or graphic) for each content unit that is to be communicated to the user. In particular, it will also be able, to opt for table and graphic presentations as supportive or illustrative presentations of the information communicated in the textual mode. The mode selection component will be rule-based. UPF will develop the rules and validate them with HSY.

For (i.2), two alternative variants are planned: a rule-based fallback variant and a machine learning-based variant. This is because of the pioneering nature of the second variant and the current lack of resources needed for it (which will thus to be acquired in the course of the project).

The rule-based variant will use discourse schemata that will need to be compiled for the major types of the bulletins to be generated by PESCaDO. The results of this task will be validated by HSY and FMI. The discourse schemata will be enriched by discourse relations as known from the Rhetorical Structure Theory (RST).

The ML-based variant will use either incremental interactive learning or learning from a previously annotated corpus (again, with a possible interaction with the user) – which implies two tasks: (1) compilation of a learning parameter set; (2) corpus annotation. It also presupposes the availability of a large and variant enough corpus. The development of the ML-based discourse structure planning is thus preceded by an assessment phase in which the ML-strategy is decided upon.

3.6.2 Information generation

The main components of the information generation submodule are: (ii.1) text generation component, (ii.2) graphic generation component, (ii.3) table generation component. The

focus of the R&D work will be on (ii.1); for (ii.2) and (ii.3), it is planned to use off-the-shelf components.

As already for (i.2), for (ii.1), two alternative variants are projected, a rule-based and a machine-learning based variant. For the rule-based variant, lexical and grammatical resources will be developed for English, Finnish and Swedish following the well-established methodology at UPF. Seed resources are already available for Finnish and English. For Swedish, the resources will be developed from the scratch.

The machine learning-based variant will be based on the stochastic Support Vector Machine generator currently under development at UPF. A prototypical operational implementation of this generator is already available and has been tested with very positive results on Chinese, English, German, and Spanish. For the SVM-generator, a substantial effort must be inverted into collection and multilevel annotation of corpora.

3.6.3 Dependencies

Closely intertwined with the discourse planning submodule is the decision module, which realizes the selection of the content that is to be communicated to the user as a decision support measure, i.e., which serves as input to the discourse planning submodule. WP 7 will be involved in the development of the content selection algorithms.

The other module, with which both submodules of the information production module will interact, is the user interface / visualization module: the discourse structure planning submodule will use it during the learning (planning) procedure, while the information generation submodule will use it for the communication of its output to the user.

3.6.4 Activity Table and Workload

In what follows, we specify the workload distribution across the development of the individual submodules and components (in PMs) and summarize the external and internal resources needed by these submodules / components.

Activity	Subactivity	Description	Dependencies	Work load
Development of the mode selection component		Algorithm for selection of the mode (text, graphic, table) for communicating a content unit to the user, depending on the nature of the content. Likely to be based on a content type – mode lookup table. Interacts with the discourse structure planning submodule.	Input structure specification for the 3 mode generators Output structure specification of the content selection module. Synchronization mechanisms for interaction with the discourse structure planning module	3
Development of the Discourse Structure Planning submodule	Discourse schema determination	Compilation of a number of discourse schemata that determine which content is to be communicated in which order.	Requires either a sufficiently large corpus or a consensus with experts on the schemata defined in a desk study	12
	Enrichment of the schemata with discourse relations	Adding discourse relations of the type ELABORATION, CAUSE, PURPOSE, etc. between the elements of the schemata. Several alternative	Availability of the discourse schemata.	

		(conditioned) relations may hold between two elements!		
	Realization of the schema-based planning component	Algorithm which (i) given an input content structure produced by the content selection module selects the appropriate schema; (ii) instantiates the schema with the received input structure; (iii) interacts with the mode selection component.	Output structure specification of the content selection module. Availability of the schemata, enriched by discourse relations. Synchronization mechanisms for interaction with the mode selection module	
	Choice of the adequate ML-based strategy	Desk study and experiments for the selection of the ML strategy to be adopted for the realization of the task. Formal description of the ML-strategy.		
	Determination of the parameter set for the ML strategy	Experiment based choice of parameters that reflect best the instantiation of discourse structures.		
	Realization of the ML-based strategy	Depends on the ML strategy chosen.	In the case of incremental interactive learning, presupposes a prior realization of the user dialogue component.	
Development of the Information Generation submodule	Selection and adapting of the graphic generator		The availability of a freely usable suitable graphic generator. Input facilitated by the discourse planning submodule in an adequate format	1
	Selection and adapting of the table generator		The availability of a freely usable suitable table generator. Input facilitated by the discourse planning submodule in an adequate format.	1
	Resource development for rule-based generation of Finnish			5
	Resource development for rule-based generation of English			
	Resource development for rule-based generation of Swedish			
	Finnish corpus acquisition and annotation			10
	English corpus acquisition and annotation			

	Swedish corpus acquisition and annotation			
	Improvement of the stochastic (SVM-based) generator			2
Development of the Content Selection Module				10

3.6.5 Timeline and interdependencies

Tasks	Interdependency	1st year 2010												2nd year 2011												3rd year 2012											
7.1 Empirical study of the material	WP4																																				
7.2 Content selection module (in connection with WP 5)	WP5																																				
7.2 Discourse structure planning component																																					
7.2.1 Discourse structure planning: discourse schema determination																																					
7.2.2 Discourse structure planning: enrichment of the schemata with discourse relations																																					
7.2.3 Realization of the schema-based planning component																																					
7.2.4 Choice of the adequate ML-based strategy																																					
7.2.5 Determination of the parameter set for the ML strategy																																					
7.2.6 Realization of the ML-based strategy																																					
7.3 Information Generation component																																					
7.3.1 Selection and adapting of the graphic generator	WP8																																				
7.3.2 Selection and adapting of the table generator	WP8																																				
7.3.3 Resource development for rule-based generation of Finnish																																					
7.3.4 Resource development for rule-based generation of English																																					
7.3.5 Resource development for rule-based generation of Swedish																																					
7.3.6 Finnish corpus acquisition and annotation																																					
7.3.7 English corpus acquisition and annotation																																					
7.3.8 Swedish corpus acquisition and annotation																																					
7.3.9 Realization of the stochastic (SVM-based) generator																																					

3.6.6 Critical factors

The critical factors for the development of the Information Production Module are: 1. the availability of corpora for training of machine-learning based algorithms; 2. the availability of an experimental knowledge base at an early stage of the project.

3.7 System Development

The development of the PESCaDO platform builds upon the modules described in the previous sections. The timing of the work and status of the demonstrator are completely decided by the progress of the individual modules and the timing of the system development work will strictly follow the original work plan.

The major check-points of the system development are marked by the milestones of the project which are described in more detail in the next section.

3.7.1 Activity table and workload

Milestone Name	Month	Modules	Sub-Modules	Man Months	Module Description	Modules required from Partners
MS2:Operational System Platform Setup	12	Architecture Setup		5	Operational System Platform for first proof-of-concept with local content and data repository set up.	Skeletons of each module : 1. discovery and targeted environmental node search, 2. service confidence assessment, 3. service node orchestration, 4. ontology web search, 5. ontology alignment and learning, 6. content distillation, 7. decision support, 8. visualization and user interaction, 9. information generation)
			Workflow Component		Workflow for a single Use Case	Use Case Specification
			User Interface		Simple UI which can demonstrate Input and Output for a single Use Case	

MS3:First Prototype	18	Workbench V1		5	Operational and extensible service platform (PESCaDO workbench) into which the components and services of the research-oriented WPs may be plugged in. Integrated modules that communicate with each other via web-services and standard protocols.	Initial versions of the individual modules
MS4:Second Prototype	26	Workbench V2		5	Extended operational service platform with knowledge and data repositories populated with the content of a representative number and variety of environmental areas and material for use cases.	Full prototype versions of the individual modules
MS5:Final System	36	Final Demonstrator		5	Operational Service Platform with full functionality of all modules and of the PESCaDO-service as an integrated system.	

4 Timing, synchronization and risk management

4.1 The development cycle and user cases

The overall strategy for the development of the PESCaDO platform is to start with simple scenarios and proceed stepwise towards the final system. The detailed resource allocations and work plans of the individual modules presented in the previous section will follow the general cycle: after the completion of each prototype version (timing defined by Milestones) of the PESCaDO-system, the functioning of the prototype is assessed, the needs for further development are identified and agreed upon, a detailed work plan for the next prototype version is finalized and the development cycle towards the next prototype is initiated. One important predefined concept for the different prototypes is the definition of the practical user cases each prototype is serving.

4.1.1 Pilot User Case 1

The range of users addressed in the first scenario covers citizens with no professional background on environmental services or clients of the services offered by HSY for the Helsinki Metropolitan Area (HMA). The citizens in Use Case 1 (UC1) are also health-conscious and outdoor activity engaged general public as well as traffic participants.

The principal duties of HSY comprise the waterworks, i.e. the production of drinking water and treatment of waste water, as well as waste management and air quality management for its four member municipalities (Helsinki, Espoo, Kauniainen and Vantaa). HSY offers air quality (AQ), and waste management information services to the citizens of the member municipalities. The population of the HMA-area is 1 million inhabitants. The current services are contracted by a number of different institutions and regularly consulted by citizens.

UC1 helps the citizens with planning a journey. It comprises issues affecting the decision of journey, such asthe weather, air quality, pollen and traffic aspects. These combined with the traffic situation make the journey fluent and minimize the exposure. These issues are useful both in routine journey planning and during air quality episodes when traffic is restricted in the city centre. In the pilot version the approach is simple and gives an answer to the question “Are there any health issues if I travel tomorrow from X to Y?”. The pilot UC1 is represented in Figure 2.

During the second SW development cycle UC1 will be extendedform passive to active – for example the PESCaDO system produces and sends alerts of issues the user has prioritized and gives suggestions of appropriate actions.

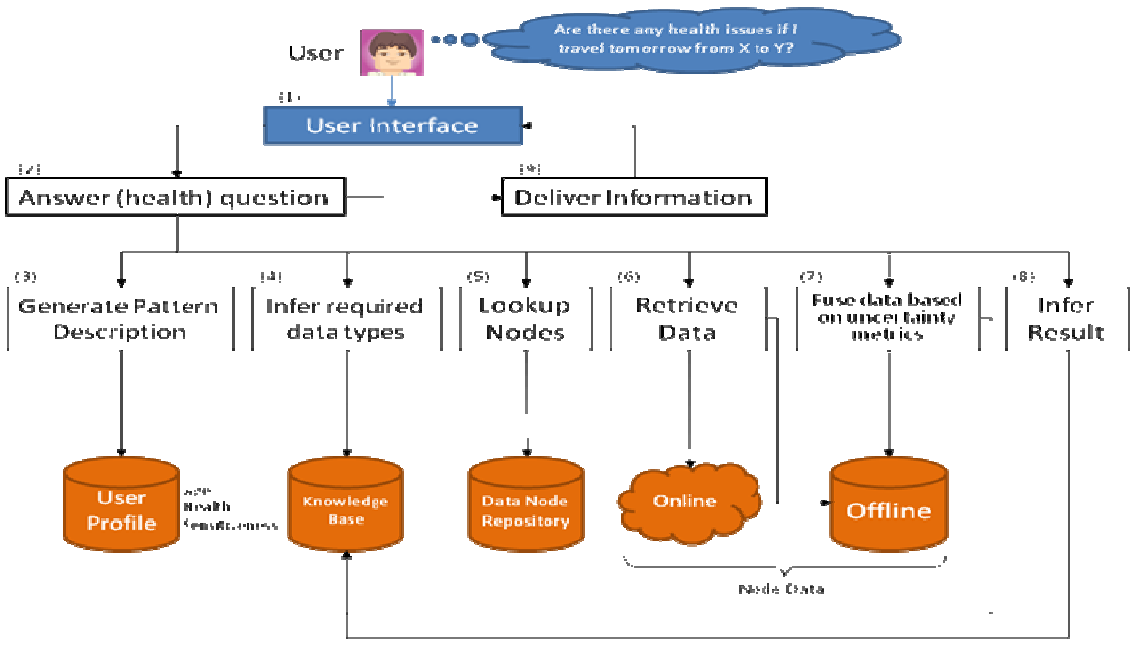


Figure 2: Pescado demonstrator for the first pilot use case

4.1.2 Pilot User Case 2

The users addressed in UC2 are experts and administrative clients of the services offered by HSY for the Helsinki Metropolitan Area (HMA). It concentrates on (administrative) decision support by combining information from different fields of activities. In UC2 the decision support will draw upon observed and forecasted environmental (AQ and weather) data, administrative action plans, information of other administrators etc.

In the final scenario UC2 will be extended to alerts and suggestions of appropriate actions to a group of authorities when certain predefined thresholds are exceeded.

4.2 Major milestones

4.2.1 MS1: Start of the project work; establishing of the project identity

Month 6 (30/6/2010)

a. Delivery of the Roadmap for the development of the PESCaDO-service, which will include the description of the scope and measurable goal of each major task and subtask in each WP, the research strategies to be adopted, stepping stones, evaluation strategies, metrics, fall back strategies, etc.
b. Delivery of the Project website and flyer
c. Delivery of competency questions verified by knowledge experts.

4.2.2 MS2 : Set-up of the operational system platform

Month 12 (31/12/2010)

Completion of the setup of the computational infrastructure of the PESCaDO system**:**

a. Presentation of the completion of the data and semantics representation framework evaluation, demonstration of the first proof-of-concept adaption and of the first proof-of-concept local content and data repository set up.
b. Demonstration of the availability of the skeletons of each module (1. discovery and targeted environmental node search, 2. service confidence assessment, 3. service node orchestration, 4. ontology web search, 5. ontology alignment and learning, 6. content distillation, 7. decision support, 8. visualization and user interaction, 9. information generation) and proof-of-concept implementations and their demonstration on restricted knowledge and data repositories, possibly with dummy data.
c. Demonstration of the setup of the communication and data infrastructure of the PESCaDO platform.

4.2.3 MS3 : 1st Prototype including the first realisation of UC1.

Month 18 (30/6/2011)

Completion of the first SW development cycle of the project:

a. Demonstration that the data and knowledge representation infrastructure is in place and capable to demonstrate the functionality of the modules that operate on it.
b. Demonstration of the initial implementation of the initial versions of the individual modules (1. discovery and targeted environmental node search, 2. service confidence assessment, 3. service node orchestration, 4. ontology web search, 5. ontology alignment and learning, 6. content distillation, 7. decision support, 8. visualization and user-interaction, 9. information generation).
d. Demonstration that all modules are integrated and communicate with each other via web-services and standard protocols.
e. Presentation of an evaluation report on the first user trials with all technologies.

4.2.4 MS4 : 2nd Prototype including first demonstration of UC2 and full implementation of UC1.

Month 26 (28/2/2012)

Completion of the second SW development cycle of the project:

a. Demonstration that the knowledge and data repositories populated with the content of a representative number and variety of environmental areas and material for use cases.
b. Demonstration of the availability of the implementation of full prototype versions of the individual modules (1. discovery and targeted environmental node search, 2. service confidence assessment, 3. service node orchestration, 4. ontology web search, 5. ontology alignment and learning, 6. content distillation, 7. decision support, 8. visualization and user-interaction, 9. information generation).
c. Delivery of a report on the performance of extensive user trials with the user-oriented technologies.

d. Delivery of a report on the first round of formal and user-oriented evaluation of all modules.

4.2.5 MS5 : Final System including the full implementation of UC 1 and UC 2

Month 36 (31/12/2012)

Completion of the 3rd SW development cycle and achievement of the research and technological objectives of the project:

a. Demonstration of the full functionality of all modules and of the PESCaDO-service as an integrated system.
b. Delivery of a CD with the documented source code of the programmes selected by the Consortium for public distribution, a demonstration image of the whole service and of the individual modules; manual for the use of the PESCaDO-service and its individual modules.
c. A detailed evaluation protocol of the individual modules as well as of the PESCaDO-service as a whole by internal reviewers and by the members of the User Group.
e. Delivery of a peer-review report of the demonstration of the showcase.
f. Project completion with delivery of an elaborated service model; a modified final Showcase (incorporating improvements based on evaluation and testing); Evaluation and Quality Assurance reports, as well as the public final report.
g. Presentation of a detailed Business Plan (including the introduction of the service to potential customer groups of the service).

4.3 Risk management

The research and development related activities are divided into research activities, technological activities and validation, and dissemination activities.

For a timely identification of risks in the course of any of these activities that are liable to endanger its objectives, the Project Coordinator will follow the Prince2 (http://www.prince-officialsite.com/) methodology, which has a well-established risk management procedure involving the construction of a Risk Issue log for the project as a whole. The Risk Issue log identifies risk situations at two levels: at WP and task level and at the project level. Entries in the log are assigned tolerances and thresholds as well as contingency plans. Risks are evaluated according to their impact and probability. Those with medium to high impact or probability are flagged throughout the whole duration of the project. A second table, the Issue Log that identifies potential or current shortcomings and dangers is maintained and supervised by the PC. The Issue Log receives input through three channels: direct communication with partners, an application embedded in DRUPAL (the web-based collaborative space) which was developed by BM to allow partners to maintain and control their list of issues, and by the PC himself. These issues are analysed by the PC and fed into the Risk Issue log if they become sufficiently serious to need management action (following the principle of “management by exception”).

The PC, together with the WP, leaders will identify the project tolerances and its contingency plans at the outset of the project so that they can be used when assessing the new risks identified. The tolerances and contingency plans will be updated throughout the project as needed.

In case of an imminent risk of a considerable delay or failure in a specific task, the Consortium decision making structure and the instruments provided by the Consortium Agreement will be activated to make the necessary corrections such as increase resources on a particular action, or to decide on any remedial measures to be taken.

5 Summary

The document presents the roadmap for PESCaDO platform development illustrating the progressively increasing functionality of the individual modules as well as of the platform as a whole. It also details the temporal and functional synchronization of the individual modules to ensure this functionality and the timeline and resources that will be needed to achieve this functionality. Together with the original, more general work plan the roadmap will assist in following and controlling the project progress during the development stages of the demonstrator and together especially with deliverable D1.1 will greatly assist in observing and reacting to any problems or deviations from original plan in the final demonstrator development process.