Project no. FP6-028038

Palette

Pedagogically sustained Adaptive LEarning Through the exploitation of Tacit and Explicit knowledge

Instrument: Integrated Project

Thematic Priority: Technology-enhanced learning

D.EVA.05
PALETTE evaluation framework and instruments

Due date of deliverable: 31 July 2008
Actual submission date: 12 September 2008

Start date of project: 1 February 2006   Duration: 36 months
Organisation name of lead contractor for this deliverable: UNIFR

Project co-funded by the European Commission within the Sixth Framework Programme		
Dissemination Level		
R	Public	**PU**

Keyword List: evaluation, usability, expertise, communities of practice, social practice, projects, …
Responsible Partner: Bernadette Charlier, UNIFR

MODIFICATION CONTROL			
Version	Date	Status	Modifications made by
1 - 6	16-07-2008	Drafts	Alan McCluskey
7	31-07-2008	Draft	Bernadette Charlier
8	11-08-2008	Draft	Alan McCluskey
9	14-08-2008	Draft	Murray Saunders
10	19-08-2008	Draft	Alan McCluskey
11	26-08-2008	Draft	Bernadette Charlier
12	27-08-2008	Draft	Alan McCluskey, Murray Saunders
13	09-09-2008	Final	Alan McCluskey

Deliverable manager
- Alan McCluskey, UNIFR

List of Contributors
- Alan McCluskey, UNIFR
- Murray Saunders, CSET
- Bernadette Charlier, UNIFR
- Joel Bonamy, GATE-CNRS
- Frédéric Vermeulin, GATE-CNRS
- Amaury Daele, UNIFR
- Annick Rossier, UNIFR
- Nikos Karacapilidis, CTI
- Denis Gillet, EPFL

List of Evaluators
- Amaury Daele, UNIFR
- Alain Giboin, INRIA

Executive Summary

This report sets out to improve the understanding of evaluation taking place in the complex context of a large multicultural, trans-disciplinary project like PALETTE. To embrace the wide variety of types of evaluation taking place, this report goes beyond conventional visions of evaluation as the province of professional evaluators to include moments of evaluation embedded in other activities carried out by what might be called lay evaluators. Evaluation is seen as generating new knowledge or revealing existing knowledge that can be used by participants in their professional practices within the project. Evaluation is seen as seeking to help diminish the complexity of situations so as to facilitate decision-making and action. This report presents a number of criteria and conceptual frameworks that can help depict moments of evaluation. The former include embeddedness, explicitness, formalness, expertise as compared to the lay approach and finally hierarchy as opposed to participation. The latter include complexity and provisional stabilities, entropy and finally communities of practice. Based on a number of examples of moments of evaluation found in PALETTE, these criteria and frameworks, amongst other things, are used to explore certain aspects of evaluation: its multifaceted nature; the relationship between the structure of evaluation and the action of carrying it out; the question of reification both of the processes and results of evaluation; and finally the challenge of the relationship between espoused and actual reality in terms of evaluation. The report goes on to evoke a number of possible ways forward that require further work including the differences in perspective at work in complex projects, the cohabitation of the evaluation expert and the layperson carrying out evaluative moments; evaluation as a means of handling complexity and facilitating decision-making; the concept of 'usability' as a lever for improvement; the need for go-betweens between communities of practice to facilitate innovation; and finally a possible way of handling the gap between espoused reality and 'actual' reality. The report concludes with a list of fifteen points as possible 'instruments' to "take to work". The annexes include brief descriptions of the moments of evaluation referred to in the report.

Note: The title of this report was fixed by the third Implementation Plan (IP3) and does not take into consideration the evolution of thinking that took place during the preparation and writing of this document.

Table of contents

1. Introduction

Aims

The aims of this text are to improve understanding of the role and place of evaluation in complex projects like PALETTE and, more widely, to contribute to research about evaluation, extending the notion of evaluation to include more informal and more implicit lay[1] forms. Such a reappraisal of the nature of evaluation, opens the way for clarification of the role of dedicated evaluation work-packages and the re-consideration of the articulation between evaluation and decision-making. This orientation is an acknowledgement that there are instances of what we might call 'evaluatively motivated' practices that are embedded in other practices associated with a project like PALETTE. It seeks to broaden what might conventionally be understood as **evaluation** by introducing the idea of **evaluative moments**, a concept that captures both the idea of embedded evaluative processes as well as more conventional or expert driven evaluations. In order to elaborate this 'reconstructed' notion of evaluation, we depict the way evaluative practices appear within the project and extend this analysis towards some more transferable learning points.

What are evaluative moments?

What do we mean by 'evaluative moments'? The phrase was coined here first and foremost to cover a set of widely found practices that retained some of the characteristics of evaluative practices as recognised by experts in evaluation without necessarily being seen as being evaluation by those carrying them out (or by those people who are professional evaluators). One of the essential characteristics of evaluation is that it generates new knowledge (or reveals or makes apparent existing knowledge or 'hidden' knowledge) that can be used by participants or others. In evaluative moments, this generation of knowledge may well not be seen explicitly as knowledge or learning by those undertaking it, just as they may not see it as evaluation. But it serves a purpose in a working process and uses some of the tools and process also used by evaluation. Evaluative moments generally entail a number of related phases that might include the negotiation of work to be done with actors concerned, the process design and the creation of tools, the collection or generation of data, the analysis and/or reorganisation of that data and the integration of this (new) knowledge in a process.

Why evaluative moments: the story of a shift in perspective

Initially, the framework behind this document was set by the third iteration of the PALETTE Implementation Plan (IP3) that says of this report: "*On the basis of the PALETTE evaluation toolset (D. EVA. 02) and on existing evaluative practice within the project, it will fill in gaps and develop an 'enriched evaluation approach' using a collaborative process with other work-packages.*"

Two deliverables fixed the framework for evaluation in Palette: D.EVA.01 'A framework plan for the evaluation and depiction PALETTE processes and outcomes' and D.EVA.02 'The Palette Evaluation Toolset'. The deliverable mentioned in IP3, D.EVA.02, aimed to elaborate "*the framework for the validation and evaluation of PALETTE services and scenarios.*"[2] The toolset was to include "*the definition of criteria for the evaluation of the adaptability, accessibility and acceptability of the PALETTE services...*"[3] This was done in collaboration with the work package for participatory design (WP1), which then went on to use it extensively in its work. The toolset was "*essentially oriented towards testing and validation of scenarios and services*"[4] in the context of work between developers and communities of practice. The toolset also explicitly excluded from its framework two 'levels' of evaluation: formative evaluation of the project itself and the generic processes of the Participative Design Methodology (PDM) that were covered by the overall framework in D.EVA.01. As a result, given that the toolset was only designed to address a specific area and form of evaluation, we need to go beyond the framework fixed by IP3 for this work if we are to "*fill in gaps and develop an 'enriched evaluation approach'*".

A design feature of a project like PALETTE is that it iterates and builds knowledge as it proceeds. In this light, it may well be the case that early statements of intent or conceptual captures, change or are

---

[1] The term "lay" is used here to mean those people and activities that are not seen to be associated with recognised experts and expertise (in evaluation). The present work might be seen as exploring the widespread role of the lay in evaluation as compared with the role and power of the expert evaluator.

[2] D.EVA.02 The Palette Evaluation Toolset, February 2007, Pg. 2.

[3] Op Cit Pg. 5

[4] Op Cit Pg. 8

transformed as new knowledge is produced. In that case, the framework elaborated in D.EVA02 is transformed in this deliverable to serve as a more global and inclusive approach to evaluation. This has to do with both who is carrying out the evaluation and the impact of the evaluative practices of the professional community that that person belongs to. We will elaborate on these ideas in what follows.

It is understandable that those people who perceive themselves as evaluators and who identify with and belong to a wider community of practice of evaluators, think first and foremost of evaluation in terms of their own practice and experience of evaluation. They have a vested interest in delimiting what is perceived to be evaluation and who is seen to do it. They are the experts. In examining evaluation within the PALETTE project, however, we were frequently confronted with 'moments of evaluation' carried out by people who did not consider themselves as evaluators and often didn't think of what they were doing as evaluation. It would have been easy to dismiss these moments as not qualifying as evaluation because they didn't fit the definition, the criteria and the standards recognised by the evaluation community. The question of the limits to set on the notion of evaluation was omnipresent throughout our work. At the same time, these moments of evaluation made up a considerable part of the evaluation in PALETTE and they were particularly rich in lessons. It is for these reasons that we have chosen to go beyond formal evaluation carried out by people who clearly identify themselves as evaluators (with a culture and practice of evaluation) to the use of evaluation by what one might call the "lay" who do not see themselves as evaluators, who may not even see themselves as doing evaluation but whose practices include elements of evaluation that are often heavily embedded in other activities. In doing so, our working hypothesis has been that examining "moments of evaluation" and raising awareness about such moments could contribute to improving ways of working and help stimulate a culture of reflexive practice. Such a hypothesis sits well with one of the evaluative approaches in PALETTE: to provide provisional stabilities[5]. That is to say that evaluation, as proposed by WP6, rather than seeking to deny complexity, is seen as providing material that helps participants understand extremely complex and fast changing realities which otherwise could not be grasped.

Scene setting

The stated aim of PALETTE, a three-year R&D project funded by EU IST programme[6], is to "*facilitate and augment individual and organisational learning in Communities of Practice (CoPs)*"[7]. To achieve this overarching aim, "*an interoperable and extensible set of innovative services as well as a set of specific scenarios of use will be designed, implemented and validated in CoPs of diverse contexts.*"[8] The R&D process in PALETTE takes place in the framework of a participatory design methodology (PDM), which was set up to establish "*a good balance between technological and pedagogical experts.*"[9] The initial description of work document (DOW) goes on to explain that evaluation is integrated into the process "*to provide direct, frequent and detailed feedback.*"[10] The evaluation implicitly referred to here is not that of the 'moments of evaluation" mentioned above but rather the more formal, explicit forms of evaluation to be carried out by the evaluation experts working for the evaluation work package (WP6).

The PALETTE project is structured in nine work packages. One of these work packages (WP1) is dedicated to developing the participatory design methodology (PDM). Three work packages (WP2, WP3, WP4) are technical in nature and are centred on the creation of services for Communities of Practice (CoPs). An additional work package (WP5) is responsible for the development and integration of the services and scenarios bringing together output for WP1 and the three technical WPs as well as feedback from users in CoPs. Two work packages concern training (WP8) and dissemination (WP7). Administration is carried out by WP0, and WP6 works on the evaluation. In addition, CoPs are associated with the work done in the different work packages, especially through WP1 and WP5. Note that the roles of WP1 and WP5 have evolved during the life of the project as project partners sought to come to terms with making the PDM a reality. Below is the work package structure as described in the 2nd iteration of the implementation plan (IP2). This schema can be compared with the diagram

---

[5] D.EVA.01. A framework plan for the evaluation and depiction of PALETTE processes and outcomes, August 2006. Pg. 4

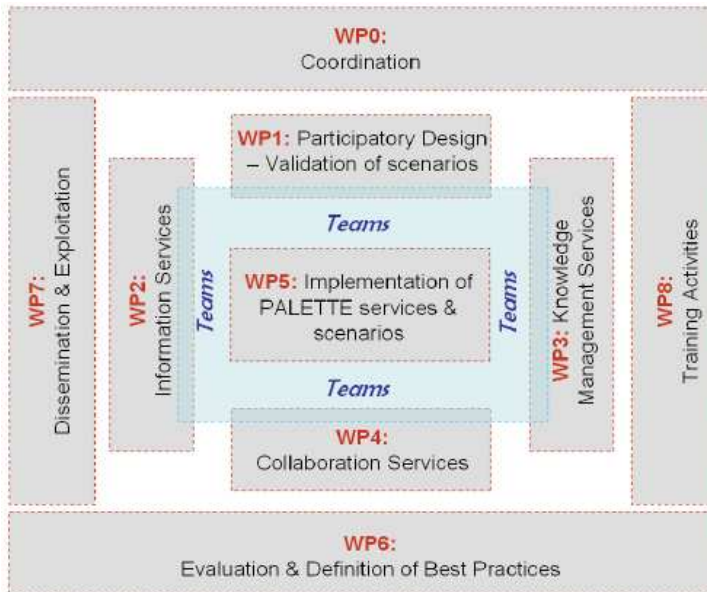[6] Sixth Framework Programme, Technology Enhanced Learning (IST-2004-2.4.10)

[7] DOW Pg. 4.

[8] Ibid.

[9] Ibid.

[10] Ibid.

depicting the working structure as envisaged at the beginning of the project (see the section on embedded and detached evaluation).

WP0: Coordination

WP7: Dissemination & Exploitation

WP2: Information Services

WP1: Participatory Design – Validation of scenarios

Teams

WP5: Implementation of PALETTE services & scenarios

Teams

Teams

WP4: Collaboration Services

WP3: Knowledge Management Services

WP8: Training Activities

WP6: Evaluation & Definition of Best Practices

An administrator and financial coordinator (AFC) and two scientific coordinators[11] (SCO, DSCO) manage the project, although, as the project description points out, the work packages "*organise and supervise the actual Integrated Project activities*"[12] within the framework of the DOW and the subsequent revised versions[13] of it. The Steering Committee composed of the project management, representatives of project partners and work package leaders is "*the most important decision taking body in the project*"[14]. It meets monthly, often at a distance. Additional critical and creative input is provided by a Scientific Advisory Board that meets once a year with Steering Committee members to discuss key issues decided on by the Scientific Coordination.

Communication between project participants flows through a number of channels. Project plenary meetings are held twice yearly and meetings within or between work packages have been frequent. All project documents are available on the project's official document server (BSCW). There are also ad hoc mailing lists for specific work packages or dedicated tasks.

Evaluation in Palette

The present document explores an extended notion of evaluation so as to take into account observed practices in all the work packages of PALETTE. These 'evaluative moments' often differ in nature from those taking place in the work package dedicated to evaluation (WP6). However, as the work of WP6 was the starting point of the current report and that work is the perceived reference in terms of evaluation in PALETTE, it is important to understand the approach adopted in WP6 as a point of comparison. To give an idea of the orientations of WP6, this section briefly points to a number of salient aspects of its approach to evaluation. Examples of evaluation from WP6 are also included amongst the moments of evaluation dealt with in this report.

The initial description of work of WP6 fixes three aims[15] that corresponded to the three levels of evaluation mentioned in the Palette Evaluation Toolset (D.EVA.02)[16]:

- Assessing the methodological approach adopted by the PALETTE project, as well as describing and operationalising this approach (adaptability, acceptability and accessibility of the open source services);

---

[11] A lead scientific coordinator and a deputy coordinator.

[12] Op Cit Pg. 38.

[13] The DOW was revised twice during the project at the same time that the Commission reviewed the work achieved: after 18 months and after 27 months. The revised implementation plans were called IP2 and IP3.

[14] Ibid.

[15] Op Cit Pg 81

[16] D.EVA.02 The Palette Evaluation Toolset, February 2007, Pg. 8.

- Providing a formative evaluation to the project at each phase (involving process and outcomes);
- Establishing how a community of practice can be supported more effectively using configurations of information, knowledge management and mediation services (evaluation for knowledge) designed with their participation and depict examples of practice that can be used as a wider community resource.

The description[17] of the evaluation work carried out by WP6 sets it in relationship with major evaluation paradigms (evaluation for knowledge, evaluation for development and evaluation for accountability) as described by such authors as Chelimsky[18]. It underlines a number of main characteristics of the evaluation: it is formative and not part of project management; it is process oriented; it provides provisional stabilities (see above) and it's collaborative.

Saunders, in preparatory work[19], explored the implications of seeing those doing evaluation as members of a community of practice (CoP) centred on evaluation[20]. In addition, D.EVA01 explicitly mentions the community of evaluators as a resource for problem solving and procedural guidance for evaluators in PALETTE[21]. Such a perspective has a number of interesting consequences. One of the central activities of such a CoP is the learning that takes place through making evaluative practice explicit. A key facet to this learning is its construction through the negotiation of meaning between CoP members and how this results in strengthening the identity of the CoP and the identification of its members with the CoP. A necessary corollary of this collective accumulation of knowledge about practice and the related strengthening of identity is that would-be evaluators need to be progressively inducted into the community.

These discussions throw in a different light the strategy behind the RUFDATA framework that was used to map out the first evaluation plan for PALETTE. It was conceived, amongst other things, as an instrument to extend awareness and capacities of those that are required to carry out evaluation but who are not necessarily recognised evaluators with the intention of inducting them into the community of practice of evaluators. As Saunders puts it: RUFDATA is "*a 'tool' for those in the process of induction.*"[22]

The thrust of the approach encouraged by RUFDATA is that it reifies some steps in the process of conscious evaluation planning by experienced evaluators and makes them available as a resource for inexperienced or novice evaluators. The critical point here however, is that the users of RUFDATA were consciously planning for an evaluation in the conventional sense. However, in exploring evaluative moments in PALETTE for the present work, it became evident that many of the people conducting those moments neither saw themselves as neither evaluators nor their work as evaluation. This would seem to suggest an alternative way of looking at evaluation as a practice of professional communities. Rather than thinking in terms of a community of practice (CoP) of evaluators, it is also possible to consider that evaluative practices belong to a wider set of practices that characterise the different professional communities taking part in PALETTE.[23] As a consequence, it might be interesting to explore if those practices differ from one community to another. For example, the types of evaluation undertaken by developers may well be quite different from those constructed by educational researchers. In addition, this perspective raises the question of the pertinence of the "inductive process" related to evaluative categories of consideration envisaged by RUFDATA when those activities are implicit and strongly embedded by people who neither see themselves as evaluators nor their work as evaluation. However, as the next section suggests, RUFDATA can be

---

[17] D.EVA.01. A framework plan for the evaluation and depiction of PALETTE processes and outcomes, August 2006. Pg. 4

[18] Chelimsky E. & Shadish W. R., Evaluation for the 21st Century: A Handbook, Sage Publications, Inc, 1997.

[19] Murray Saunders (2000) 'Beginning an Evaluation with RUFDATA: Theorising a Practical Approach to Evaluation Planning' from Evaluation Vol 6(1): 7-21

[20] See Wenger E., Communities of Practice: Learning, Meaning, and Identity, New Ed., Cambridge University Press, 1999 and Lave J. & Wenger E., Situated Learning: Legitimate Peripheral Participation, Cambridge University Press, 1991.

[21] PALETTE (2006) 'D.EVA.01 A framework for the evaluation and depiction of PALETTE processes and outcomes', PALETTE.

[22] Saunders M., "Beginning an evaluation with RUFDATA: theorising a practical approach to evaluation planning," Evaluation, Volume 6, no.1 : 7-21.

[23] Such a perspective (of multiple communities, each with its own practice of evaluation) might also provide a more satisfactory framework for understanding the 'community' of evaluators within which many differing, if not diverging, practices cohabit.

reconstructed to form an interrogatory tool to bring into relief evaluative moments in wider areas of practice.

Participation, communities, peer learning and design

In this short section, we continue to set the scene in PALETTE by briefly describing the central pillar of the project: the Participative Design Methodology (PDM) and how this relates to the notion of community of practice, to peer learning and to evaluation.

According to the PALETTE Description of Work (DOW), communities of practice (CoPs) are "*effective environments to support learning by professionals, organisations and educational institutions.*[24]" This learning is often collaborative in nature, where "*members learn from each other by making their knowledge and practices explicit, sharing them with peers and reflecting on them.*[25]" Four challenges related to learning in CoPs are singled out for attention in PALETTE:

- Expressing, representing and sharing practices and authentic problems.
- Debating and reflecting about practices and about life of the CoP.
- Developing, reifying and exploiting knowledge inside and outside of the CoP.
- Facilitating engagement, participation and learning.

It is interesting to note that evaluation is not explicitly mentioned amongst these challenges. This may not imply that 'evaluation' is absent, but rather that those aspects of these challenges that are related to evaluation (for example in reflecting on practices or developing knowledge or facilitating learning) are highly integrated and not seen to be evaluation.

In addressing these challenges, "*the points of view of both developers and CoP members are closely interrelated through a distributed participatory design methodology (PDM) in order to develop technological services that could support CoP needs for information sharing, knowledge management and collaboration.*[26]" The PDM takes the form of an iterative process in which CoPs and developers of services work together, identifying CoP needs and activities, developing technological and learning services based on these needs and activities, designing scenarios of use for these services and organising activities through which CoPs could participate in the design of the services. Evaluation is then necessarily deeply interwoven into this collective learning, design and development process. The work package dedicated to the participative design methodology (WP1) is the centre of the most extremely elaborate forms of evaluation in the project.

The method employed

The method adopted here is a grounded approach[27] in which the concept of 'evaluative moments' and ways of depicting and understanding those moments emerged from the collection and analysis of data about practices related to evaluation within the context of a large scale European funded R&D project.

Material for this report was first collected from project documents. Structuring this data led to a number of observations about the nature of evaluative practices in the project. These observations were used to formulate a set of strategic questions that guided subsequent interviews of a number of key figures from the various work packages. These questions were built on and extended the existing RUFDATA framework (discussed briefly in the next section), adding such aspects as the process of creation of instruments used, the integration of the evaluation process in the rest of the activities of the project, the implication and engagement of other actors in the process of evaluation and the form taken by the results.

An initial overview of the evaluative moments on a work package by work package basis was drafted. This was used to identify a series of axes that could be used to portray salient aspects of such moments of evaluation. These were tested against the cases studied. Subsequently, a number of examples of moments of evaluation were selected to illustrate the aspects in the present document. Brief descriptions of these selected moments can be found in the annexes. The variables elaborated to depict moments of evaluation are described in the following section.

---

[24] DOW, Pg 4

[25] Ibid

[26] Daele A et aliud, Participatory Design for Developing Instruments for and with Communities of Practice: a Case Study, CHI 2008, April 05, 2008 Florence, Italy.

[27] Charmaz, K. (2006). *Constructing Grounded Theory. A practical guide through qualitative analysis*. London: SAGE

2. Depicting moments of evaluation

The task of this section is to bring together a series of ways of approaching moments of evaluation such as to elucidate how these moments contribute to, or occasionally hinder, project processes and learning. We developed some units of analysis (that is ways in which evaluative moments might be understood and explained: for example, in terms of the degree of formalness or explicitness). This was in itself a learning process typical of a grounded approach, so we begin here by providing a short description of part of that process before going on to explore the different analytical frames from which we planned to approach the selected moments of evaluation from the PALETTE project in the following section.

At an early stage of this work, it was planned to gather data and analyse it using the RUFDATA framework. That framework, which was used to structure the initial work of WP6, is made up of seven questions concerning the reasons for the evaluation, its focus points and its nature (data gathered, audience, timing, and who conducts the evaluation). A complete list of the RUFDATA questions can be found in the annexes at the end of this document. As mentioned above, RUFDATA[28] was conceived initially as a framework to help plan evaluation especially on the part of practitioners seeking to carry out self-evaluation. RUFDATA was designed for forms of evaluation that are formally structured and explicit or seek to become so. However, as we note above, many of the evaluative moments studied in PALETTE were neither explicit nor necessarily formal and they were often not even perceived as evaluation. As a result, additional parameters were necessary to help describe and understand these moments of evaluation. Our first approach concerned the design and intentions of the evaluation. We then looked at a series of 'axes' (for example, between formal and informal) that formed our analytical frames and help to 'position' the moments of evaluation. And finally, as we studied the cases in more depth a number of conceptual frameworks helped us further clarify the moments of evaluation.

The remainder of this section describes those parameters and conceptual frameworks that can be used to help depict moments of evaluation.

Types of evaluation in terms of intention and design choices

One way of looking at evaluative activities is to consider the intentions that underlie those activities. Much evaluation has to do with **piloting** the wider context or helping make choices about the orientation of a particular activity (see the work of the Scientific Advisory Board or that on the revision of the Implementation Plan, for example). The intention may be **formative**, driven by a quest on the part of participants to learn or to change their ways of working (for example, the so-called Lyon evaluation[29] about formative evaluation as with much of the work done in the evaluation work package). That quest for knowledge may be motivated on a broader level by a desire to contribute **knowledge for science** (the current report is an example of such work). Some forms of evaluation are anchored in the intention to **empower**, enabling people to decide for themselves (the work on formative evaluation might be considered empowering in that it potentially fosters reflexive practice). In a more political and policy-related direction, the intention may be that of **justification** or legitimisation in which evidence is provided to justify a particular course of action. The intention may also be to provide **accountability** where conformance is tested against pre-determined expectations often stemming from an external authority.

The characteristics of evaluative practices in terms of intention will naturally affect the design of evaluation. Other choices also affect that design. One such choice is the relationship to **change**: to what extent and in what ways does the evaluation process intentionally or incidentally affect the context in which it takes places (for example, the evaluation of ways of communicating – see Annexes for details – might modify the way people work within the project). Another design choice concerns the degree of **integration** of the evaluation moment as part of a wider process (the work of the Trials – see Annexes – involved results that needed to be integrated in a number of other work packages). The level of **engagement** of actors in the evaluation process is an additional example of a parameter

---

[28] Saunders M., Beginning an evaluation with RUFDATA: theorising a practical approach to evaluation planning, In Evaluation Volume 6 no. 1 pp 7-21

[29] This evaluation took place during the Palette Plenary Meeting in Lyon in December 2007 during which Steering Committee members were interviewed (see Annexes for details).

of the design of the evaluation moment (engaging a wide range of actors in elaborating the business plan was part of the work on WP8 – see Annexes for details).

Five axes

The initial study of the evaluative moments in PALETTE suggested five[30] axes along which each moment could be positioned. They were particularly interesting not so much because they enable a clearer depiction of evaluative moments, which they did, but because when taken together they promised new insights about the nature of such evaluation and its role in projects. Having said that, as will be seen when we examine more closely examples of moments of evaluation from PALETTE later in this text, these apparently clear distinctions (for example, between formal and informal) are not at all so clear-cut in reality.

The use of the metaphor of an axis or a continuum might give the impression that each moment is situated at unique coordinates along these axes. As can be seen in what follows, this might not necessarily be the case. For example, the espoused design might be formal, whereas the reality of the evaluation might be more informal. For this reason, a graphical representation in terms of a series of crossing axes might prove to be complex and, in addition, the positioning of the axes with respect to each other might suggest relationships that don't exist.

For the time being, however, lets describe these 'axes' as they still prove to be rich in providing insights.

These five axes are:
- Embedded – external
- Explicit – implicit
- Formal – informal
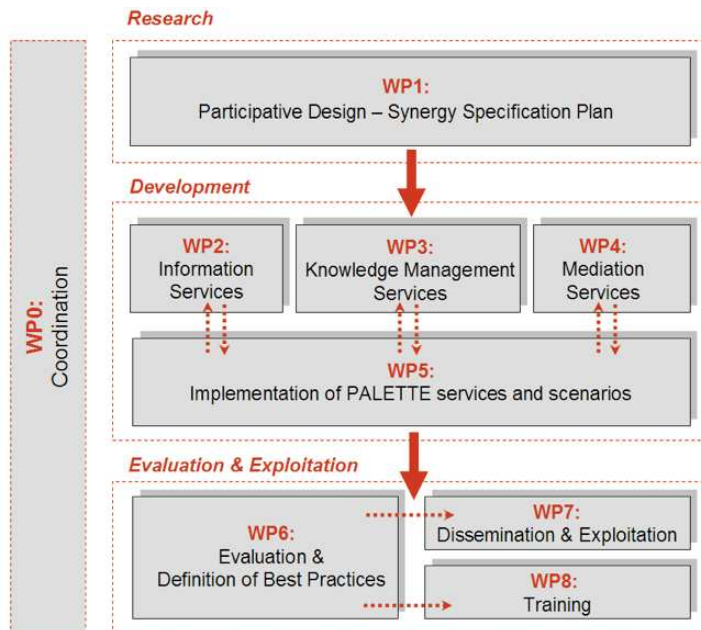- Expert – lay
- Hierarchical – Participative

Embedded – external

One of the aspects that characterises many of the moments of evaluation found in the PALETTE project is the level of embeddedness in the project or in the work package processes. The word "embeddedness" may be understood in several ways here.

On the one hand, there is the extent to which the evaluative activity is perceived to be independent from the other activities in terms of the degree of liberty to go beyond or to challenge existing ways of working or pre-conceived ideas about the functioning or the expected outcomes of the project or work at hand. Such a facet of embedment is related to the attitude to and the possibilities for change. This relationship can be somewhat complex. Let's take the example of the authentification of scenarios in the participative design work package (WP1) (see annex for details). Within the process of checking the scenarios as appropriate for CoPs, it was normal for parts of the scenario to be questioned, as that was an integral part of the work. It would not have been appropriate; however, to challenge the idea of using scenarios, as such a level of abstraction was tacitly not part of the task even if the experience of the task might hypothetically have led to that conclusion. In other words, in certain types of evaluation, the form of the evaluation and possibly the assumptions of the evaluation are givens that cannot be challenged within the context of that evaluation. The current work on evaluation is a counter example in which the initial assumptions of the evaluation are being explicitly deconstructed.

On the other hand, embeddedness may be considered in terms of the interrelatedness and the interdependence between the evaluative activities and the other activities they are connected with. To what extent can the evaluative activity be separated from related activities? To what extent is it standalone? To give an example, if you remove from the Participatory Design Methodology (PDM) process the authentification of the scenarios developed for and by CoPs in WP1, the process of creation of services and scenarios breaks down. In comparison, if you take away the work on perceptions of formative evaluation in WP6, the project will continue to function, even if the level of understanding of project processes may well be diminished and the project management may possibly be less piloted by foresight.

---

[30] There may well be others, but we have concentrated on these four.

Research

WP1:
Participative Design – Synergy Specification Plan

Development

WP0: Coordination

WP2:
Information Services

WP3:
Knowledge Management Services

WP4:
Mediation Services

WP5:
Implementation of PALETTE services and scenarios

Evaluation & Exploitation

WP6:
Evaluation & Definition of Best Practices

WP7:
Dissemination & Exploitation

WP8:
Training

The "externality" of much of the work of evaluation work package (WP6) from the everyday work of the other work packages was pointed to by a number of Steering Committee members in a study about formative evaluation in PALETTE[31]. It is interesting to note that, according to the perception of some interviewees, the type of evaluation carried out by WP6 is seen to need a level of externality to function correctly. The diagram[32] of the PALETTE Work Packages as it figures in the Description of Work (DOW) elaborated at the beginning of the project (see above) might be interpreted as graphically illustrating this 'externality'.

Explicit – implicit

The distinction between explicit and implicit lies in the extent to which the nature of an activity or an intention is openly expressed. An activity may be implicit because those people organising it take it for granted and don't feel they need to explain what they are doing or why they are doing it. The implicitness of the processes and the assumptions in the work on the future PALETTE business plan, for example, were possibly some of the reasons why participation in the process was disappointing.

This is a particularly significant axis within PALETTE because much of the learning seen to take place in CoPs consists of making tacit knowledge explicit and PALETTE itself is sometimes likened to a CoP. In the goals expressed by WP6 for evaluation in the project, making the implicit (or unexpressed) aspects of the work more explicit was seen as a way of improving evaluation practices.

Formal – informal

The word 'formal' refers here to the extent to which the evaluation activity is explicitly designed according to an evaluation approach which is carried out more or less systematically and driven by an evaluation model and a predetermined plan. There will always be a difference between the formality of the intentions underlying the design of the evaluation and the reality of its implementation. For example, the Steering Committee planned to set rules covering the review of deliverables (WP0) whereas in reality review procedures continued to depend largely on the individual choices of reviewers. Another such example would be the preparation of the review process where a series of indicators were developed by WP6 to help guide the assessment of the previous year's work, whereas, judging from the reported discussions in the Steering Committee minutes, little attention was paid to these indicators.

In comparison, an informal approach involves a less pre-determined course of action. As a possible example of a more informal approach, consider the testing of possible conferencing facilities undertaken by the SC in WP0. No structured procedure or specific instruments were planned to carry out this evaluation. In addition, the feedback from these "tests", as presented in the SC minutes was

---

[31] McCluskey A., On formative evaluation as a lever for change, PALETTE WP6, June 2008.

[32] Quoted in D.EVA.02 – The Palette Evaluation Toolset, PALETTE, 2007, Pg 4.

not formally structured or documented. It was not judged necessary to prepare a report or list outcomes and possibilities.

The Trials (see the annexes for a brief description) offer an interesting case where formality and informality co-exist. The protocol for observation and analysis of the use of PALETTE tools in given scenarios was extremely elaborate and formal, although individual mediators who had to carry out the work were free to adapt it to the context of use. The process of use to be observed, however, was full of many small informal moments of exchange and evaluation. At the same time, the reformulation and synthesis of the specific scenarios in a small number of generic scenarios was dictated by no formal procedure.
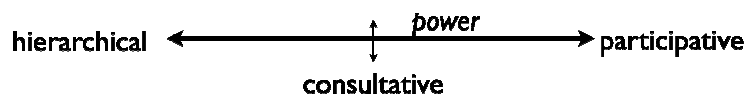
Expert – layperson

The two terms 'expert' and 'lay' both refer to knowledge or skills in a particular area, but in the former case that body of knowledge is recognised and the person is seen as an authority, whereas in the second case the knowledge is not (officially) recognised and the person is seen as not qualified (in the sense of being sanctioned by a diploma or by membership of a professional body, for example). Using the word 'lay' in no way implies a lack of seriousness or rigour. One might consider that such a distinction favours formal knowledge over tacit knowledge, although both experts and laypeople capitalise on a stock of tacit knowledge and personal experience. The distinction between the expert and the layperson also brings up the question of authority, power and influence. We will talk about power later when discussing the difference between hierarchical, consultative and participative approaches to evaluation. However, we discover, when considering PALETTE, that lay practices in evaluation are endemic in behaviours associated with informal reviews, reflections, planning and 'what we have learned' meetings, adjustments in design on the basis of 'what works' etc.

This divide between expert and lay can be found between those considered to be evaluators and as such recognised members of the community of evaluators and those people carrying out evaluation within their own practice but who are not considered to be official 'evaluators'. The divide can also be found between people taking part in the evaluation, for example between educational researchers and computer scientists on the one hand and 'users' of technology (and members of CoPs) on the other as in the case of authenticating scenarios in WP1. These differing configurations of expertise versus lay experience and the related power relationships will necessarily affect the design and the outcomes of the evaluation.

Hierarchical – participative

It is also interesting to look at these moments of evaluation in terms of involvement and the role of participants in the evaluation and how these relate to the question of power and authority. In a hierarchical approach, intentions, design and analysis are all decided elsewhere. They do not concern most participants except in that they may be required to provide work or knowledge at the appropriate moment. In a fully participative approach, participants are involved in deciding on intentions, in formulating the design of the evaluation, in carrying it out and in analysing the results. In the participative approach power is delegated more equally to participants. Somewhere between the two the consultative approach seeks the opinion of participants. While their knowledge is tapped and to a certain extent can influence what happens they are not given additional power of decision or further involved in the design or realisation of the evaluation.

These ideas spring from the work of describing the authentification process in WP1 and the respective roles of the user and the expert. As mentioned above in the discussion of lay and expert knowledge, the respective status of the layperson and the expert confers or withholds power, which can produce an unbalanced relationship between developers, researchers and users in which full participation on an equal footing is quite difficult especially as the process is largely centred on knowledge (the very place where power relationships are at play here).

Seen in the light of this discussion, the Participatory Design Methodology (PDM) in its attempts to involved users, developers and educational researchers in developing both services and their use, might be seen as a "*partial democratisation of the development process.*" Depending on the success in fully involving participants and handling the relationship between the lay and the expert, the process might be participative or consultative.

From a different perspective, it is interesting to look at evaluation in terms of its 'inclusiveness'. Saunders[33] argues for *"an 'inclusive' evaluation stance from a moral/political standpoint and from the standpoint of sound evaluation design.*[34]*"* Inclusiveness is understood to be the extent to which the 'voice' of those people or groups with a latent or explicit interest in the evaluation process is taken into consideration. The description of evaluation by WP6[35] refers to the desirability of an "inclusive" approach to evaluation. A consultative approach would suffice to guarantee inclusiveness as it is defined here.

Complexity and provisional stabilities

In addition to employing the axes mentioned above, we can use several other conceptual frameworks. One concerns complexity and is partly inspired by actor-network theory[36] (ANT) in which activities are constituted of and emerge from networks of actors (both human and material). The nature of these networks is continually emerging and can at no time be considered as fixed or certain. Such a situation would be unworkably complex if it weren't that there are 'provisional stabilities' both in actors and in activities. Law calls this 'punctualisation' where a series of actors and actions 'coalesce' to form a whole that is then perceived as a single actor, at least momentarily, thus simplifying activities and facilitating understanding. It is only when this 'actor' dysfunctions, that is to say does not behave as predicted or expected, that the 'actor' disaggregates and is seen again in its full complexity as a series of separate actors and activities in their own right. Law gives the example of a television, which, when it doesn't work, appears for what it is, a complex network of 'actors', both human and material. When dysfunction occurs, the level of perceived complexity rapidly increases threatening the functioning of the wider system by making understanding, decision-making and action more difficult if not impossible.

Evaluation could be seen, amongst other things, as contributing to diminishing complexity, albeit on a temporary basis, so as to facilitate understanding and decision-making. It is in this sense that the term 'provisional stability' is used in work on evaluation in the PALETTE evaluation work package as *"… sense-making knowledge produced by reflection on and the understanding of change, enabling choices or decisions for future action.*"[37] There is a subtle difference, however, between this sense-making work and the punctualisation pointed to by Law: the former is an explicit and voluntary act designed to diminish complexity whereas the latter is a phenomenon of spontaneous emergence that in many ways lies beyond control. And, just as the emergence of 'actors' that reduce complexity is largely unpredictable, so the diminution of complexity due to strategies of evaluation cannot be guaranteed, all the less so as the complexity of the situation increases.

As a partial illustration of the question of complexity in evaluation, let's briefly take the example of the Lyon evaluation about formative evaluation and decision-making. When asked about evaluation and project management, the answers of members of the Steering Committee pointed to many different and often diverging conceptions and opinions. The work of analysing these results consisted of reorganising them in what might be called 'constellations' of similar conceptions so as to apparently reduce the overall complexity of the situation.

Entropy

Entropy is a concept that originated in thermodynamics where it refers to the measure of unavailability of a system's energy to do work. It is often associated with a decline of order and an increase of disorder. One of the interesting aspects of complexity in relationship to entropy is that it would seem to work against the supposed irreversibility of entropy by enabling order to emerge spontaneously from complexity. The concept of entropy is also employed in information theory and sociology.

In our context, entropy provides an interesting metaphor for the loss of knowledge when it is no longer available for use by a given system or when knowledge is no longer usable because of the timing of its availability. This is particularly important in evaluations that purport to be formative. There is a tension between the rapidity of feedback required to be genuinely formative in a fast moving project

---

[33] Saunders, Murray (2006) 'The 'presence' of evaluation theory and practice in educational and social development: toward an inclusive approach', London Review of Education, 4:2, 197 - 215

[34] Op Cit Pg. 197.

[35] D.EVA.01.

[36] See in particular: John Law, 'Notes on the Theory of the Actor Network: Ordering, Strategy and Heterogeneity', published by the Centre for Science Studies, Lancaster University, Lancaster LA1 4YN, at http://www.comp.lancs.ac.uk/sociology/papers/Law-Notes-on-ANT.pdf

[37] Saunders, Murray (2006) 'The 'presence' of evaluation theory and practice in educational and social development: toward an inclusive approach', London Review of Education, 4:2, 197 - 215

and the time required to build more 'stable' knowledge about what is happening. The functioning of this metaphor depends on two ideas. First there is the idea that usefulness can be taken as a measure of the value of knowledge. Second there is the idea of availability of knowledge and that unavailability might be considered permanent. The concept of availability of knowledge, given what we know of tacit knowledge as opposed to formal knowledge or explicit knowledge or the unconscious compared with the conscious, requires a more nuanced approach.

Evaluation can be seen as making additional knowledge available for use by participants as well as indicating where useful knowledge has been lost or forgotten or ignored. The relationship between design and use in evaluation, as explored below, can be seen in terms of entropy. To what extent does the evaluation process lead to the development and use of new knowledge and thereby works against entropy? The intention of all evaluation is to generate new knowledge (or reveal or make more apparent and thus available existing or 'hidden' knowledge) that can be used by the participants. Not all knowledge that could be thus generated is seen as useful or worth the effort. Take the example of the workings of the Scientific Advisory Board (SAB) discussed below in the section about reification. It might have been useful to explicitly map out the decision process leading to the choice of discussion subjects. However nobody bothered to do so. It wasn't even a topic for consideration. To what extent was entropy at work in that potentially useful information was 'lost', or rather not found? This is both a question of evaluation design on the one hand and of use on the other with the latter seen from multiple perspectives including that of management (fixing priorities), economy (time and means available) and power structures (who decides what for whom). Entropy was certainly at work later in the process where the lack of sufficient notes on the proceedings led to a loss of knowledge generated in the work of the SAB. However, the usefulness of what was 'lost' can only be ascertained later depending on the circumstances, making decisions about what to conserve more difficult.

Communities of practice

Another conceptual perspective that is useful here is inspired by that of the community of practice[38] where developing practice by exchange and negotiation between members of the community leads both to learning and the creation of identity. Developing practice, and, as a result, learning in a community of practice depends heavily on what's is called 'reification', that is to say the transformation of knowledge into a material form, whether it be a set of notes, a web-site or a significant object for participants, for example. This transformation does not necessarily imply increased formalisation. For some, the dynamic of communities of practice is seen in terms of growth, development and survival of the community, where inducting new members is a key activity. More important for us here, is the relationship between professional practice within a group and how these practices are negotiated leading to learning and improvement. As mentioned elsewhere here, aspects of the evaluative moments can be seen as part of a set of professional practices that are cherished, maintained and defended as part of a collective and individual identity. This potential attachment to existing practices could offer resistance to change. On the other hand, increasing awareness of those moments and the related practices contributes to the process of strengthening both the knowledge and the identities of the communities concerned.

Without entering the debate about whether or not a professional community can legitimately be considered a community of practice as Wenger and others consider it, there is one point that needs clarifying. Differing and even divergent practices and related belief systems and conceptual frameworks often co-exist within professional communities. The 'discipline" of evaluation as Chelmsky and Shadish call it[39], has its own divergences. Suffice it to quote some of the section titles from Chelmsky's introduction to their book on evaluation[40]: "*Differing purposes, perspectives and methods*" or "*Differing views about use and the evaluator's role*". As a result, when considering the practices of a professional community one should not necessarily expect a unified or common approach.

---

[38] See in particular, Wenger E., Communities of practice: learning, meaning, and identity, Cambridge University Press, Cambridge, 1999.

[39] Op Cit, Preface Pg. xi.

[40] Op Cit Pgs 1 – 26.

3. Interpreting moments of evaluation

In what follows, we set out to look at a selected number of moments of evaluation from the PALETTE project. This is partly done in terms of the axes, criteria and conceptual frameworks evoked above in Section 2. As will be seen, examining the moments of evaluation also suggests new perspectives and other ways of approaching them. Brief descriptions of a selection of moments of evaluation from the Palette project can be found in the annexes.

The multifaceted nature of evaluation

If one of the functions of evaluation is to diminish complexity so as to enable or improve understanding and decision-making, evaluation itself, on closer inspection, is extremely complex and multifaceted. Arguably it might not be necessary to understand that complexity to employ those practises. As David Wood, formerly of Nottingham University, put it with humour, "thinking too much about how we walk down stairs while doing it often causes people to trip and fall". However, if we seek to improve practices, then extending our understanding of them might prove useful. What we set out to do in this section is to point to some of the aspects of the multifaceted nature of evaluation through the nature of the work carried out.

What we call here 'evaluative moments' are many things to many people. If our hypothesis that practices used in moments of evaluation are often part of wider sets of practices belonging to given professional communities, then one could expect practices to differ from one community to another. These differences could extend to the phases of the moment of evaluation. Confirming these hypotheses goes beyond the scope of the current work.

In our clarification of evaluative moments above, we mentioned a number of possible phases to the evaluation. It is interesting to examine the moments of evaluation in terms of these phases and the extent to which they are present. The phases are as follows:

1. **Negotiation of the work to be done** – It is important to note that in many of the cases presented in this report the overall framework of the work to be done has been previously negotiated with other partners and the European Commission in the process of revising the implementation plan. This framework does not go into details about the work to be done. In some cases there is very little or no further negotiation between participants of the nature of the process to be undertaken or even the object of the process. For example, in the preparation of the Scientific Advisory Board meetings the reasons for having such a Board are not discussed. Nor is the format of the meeting officially discussed. Common practices for the organisation of meetings are tacitly adopted. In addition, there would appear to be no formalised discussion of the objects of meeting (the issues to be treated) prior to it taking place. The situation was different in the preparation of the Trials (WP1 & 5). In the case of the Trials, negotiation is part of the structure of the work as much of the work is discussed and decided on in the framework of cross-disciplinary teams made up of pedagogues, developers, members of CoPs and mediators between the various groups of actors involved. These discussions take the form of online meetings, face-to-face meetings and online collaborative working (using a Wiki for example). It is interesting to note that, in order to accommodate diversity, to leave room for differing professional practices and to cater for particular local contexts, a minimum framework was decided on (for example, in the choice of activities to be observed, where four criteria were elaborated) within which actors were free to make their own choices[41].

2. **Process design and the creation of tools** – A number of the moments of evaluation examined here had some element of process design, especially those belonging to the evaluation work package (WP6). In the Lyon evaluation, for example, a set of questions was drawn up prior to the interviews as well as a summary of earlier outcomes as a stimulus for discussion. Many of the evaluative moments in the participative design work package (WP1) had elaborate and complex process design phases. The design of the Trials, for example, was a particularly elaborate process with an explicit and formal methodology that provided a general and a conceptual framework for the work as well as a detailed schedule and

---

[41] (2008), WP1 – T4B: Methodology for the observation of the Palette services for trials, (version 10 – June 2008)

methodological tools[42]. Note, as an indication of the iterative nature of this work, that the current version of that methodology is #10. Yet, as mentioned above, within that set framework actors had a great deal of liberty to design their own processes and tools. On the other hand, the less formal moments of evaluation had no formal process design and if evaluation tools were employed they were lifted from existing practices and tools. See for example the Testing of communication tools in WP0 where no explicit work was done on deciding on how such communication tools would be tested.

3. **The collection and/or generation of data** – In formal evaluation, that is to say where the form of the evaluation is decided on and laid out in advance, the collection of data is the aspect of the process that generally receives the most attention. It entails fixing the questions to be asked and delimiting the processes to be observed. It involves choosing ways and means of observing: interviews, questionnaires, video or sound recording, direct observation, monitoring computer user,…

Sometimes the 'collection' phase comes in the form of gathering feedback. For example in the Authentification of scenarios (WP1), CoP members provided feedback either in focus groups or via questionnaires about the scenarios developed on the basis of their expressed activities collected in a prior stage of evaluation.

There can be several data collection moments in a process, reflecting shifts in the focuses of the process. For example, in the Review Process (WP0), data is first collected amongst Steering Committee members to prepare the presentation for the reviewers. Data is then collected during a mock-up of the review that takes place amongst project partners that is used to revise the presentations and strategies for the review. Data is then collected during the review itself in terms of feedback and questions of the reviewers as well as the input from project partners. Finally, formal feedback comes from the reviewers report and recommendations that are then integrated into the process of redefining the implementation plan.

The Deliverable Evaluation process (WP0) is an interesting limit case where observation and analysis take place simultaneously. The evaluator reads the deliverable (observes) and at the same time identifies those problems present in the text (both observes and analyses). Note that in this case there is no structure for either the observation or the analysis, lest it be the nature and structure of the document itself.

4. **The analysis and/or re-organisation of that data** – Although the analysis of data from documents, observation, questionnaires, interviews, etc. is one of the key moments in the evaluation, the way the analysis is carried out is often largely intuitive. The detailed methodology for the Trials only very scantily covers the analysis of data, referred to as 'content analysis methods'[43]. This phase is largely left to the discretion of observers and researchers. Observing early results, a common practice for the analysis involves breaking down the research questions into smaller units and then using a grid to associate these "units" with the relevant parts of the data collected. In a second phase, which often overlaps the first, the material in the grid is re-assembled and re-organised with a view to "making sense" of the observations in the light of the research questions. This sense making entails a simplification through the 'uncovering' of patterns or structures. The Crete Interviews[44], discussed below, illustrated this process of re-assembling the data. However, as one of the researchers pointed out, the positioning of material in the analysis grid and the nature of the subsequent patterns observed depend heavily on the individual carrying out the analysis, on his or her culture and professional background and on the extent to which he or she is implied in and aware of the process being observed.

In comparison, in the Scientific Advisory Board analysis and data collection are partly simultaneous. Some of the analysis of the discussions with experts is an integral part of the discussions themselves. Analysis also takes places in subsequent informal, bilateral discussions and probably in the personal, undocumented reflection of certain actors directly involved in decision-making. Although the discussions were presented to project participants in Plenary Sessions and evoked in Steering Committee meetings, in particular prior to revising

---

[42] Op. Cit.

[43] Op. Cit. Pg. 11.

[44] These interviews took place during the Palette Meeting in Crete in Summer 2007.

the implementation plan, no formal analysis of the data took place there or in any other context.

The unanswered question here is: to what extent would the necessary investment in time, effort and involvement to carry out more structured analysis of outcomes prove to be worthwhile?

5. **The integration of this (new) knowledge in the process** – In many of the moments of evaluation examined, the integration of results into the process was an essential part of the work. And that integration necessarily entailed decision-making. It is arguably this integration that distinguishes many of the moments of evaluation from the more conventional evaluation forms. A number of professional evaluators argue that evaluation should provide data that is accessible and useful to decision-makers, but should not provide recommendations[45]. That, they say, is the role of the decision-maker. This clarification of roles is no doubt essential in conventional evaluation where the evaluator is often an external expert called in to carry out a formal evaluation. In many of the moments of evaluation studied here, however, the main actors were often those carrying out the evaluation as an integral part of their working processes. In the case of the Trails (see the annexes for a brief description), for example, representatives of developers, pedagogues and members of CoPs as well as mediators between these professional groups participated in the evaluation process, each with a vested interest in integrating the results into their work. Note that the final assessment and integration in a particular professional activity was understandably left to the professional community concerned.

The integration phase during which the outcomes of the evaluation were fed back into the core process was downplayed by the evaluation work package in PALETTE for the professional reasons given above. The outcomes of the evaluation were seen as resources for project partners[46] but the integration of those outcomes was not seen as part of the work of evaluation. This can be seen in format of the outcomes of the Crete interviews[47]. The Steering Committee was presented with a three-page document containing the answers to three questions in the form of 'quotations' assembled under a series of headings and subheadings. Such a presentation reflected the work of analysis in 're-assembling' the collected data to reduce complexity (as mentioned above) but it was also indicative of an approach that set the onus of reaching conclusions and coming to decisions on those who were recipients of the results. It is interesting to note that, on the part of some people, these results were received with incomprehension and even dissatisfaction that was echoed in the subsequent Lyon interviews[48]. In their perception of evaluation, those people expected the work of evaluation to embrace the making of recommendations but also the animation of discussions of the results in the Steering Committee. Note that, in contrast, the 'Lyon interviews' report terminates with a series of recommendations in the form of 'possible pathways forward' in response to the request of interviewees that "*WP6 accompany evaluation results with pointers indicating possible directions that the project could take in the light of the evaluation results*"[49]. The actual nature of subsequent discussions about these results and recommendations and the importance granted to the evaluation results remained the prerogative of the Steering Committee and, above all, the project management.

Structure and action

Let's turn our attention to the relationship between the structure and organisation of the evaluation activity (frameworks for action) and how it 'acts' in and with the context (action itself). Making such a distinction between structure and action might be misleading if we do not recognise that they are two facets of the same thing. Clearly the structure (and design) of the evaluation is going to have an impact on the way the evaluation acts on the context and, at the same time, if the evaluation seeks to be successful, the way it acts is going to depend on the way it is organised. As mentioned above the dysfunction of a part of an activity or actor generally leads to increase complexity, which in turn

---

[45] Saunders, M., 2006. Do We Hear the Voices?: The 'Presence' of Evaluation Theory and Practice in Social Development. Evaluation, 12(2), 251-264., Pg. 255.

[46] (2006), D.EVA.01, A framework plan for the evaluation and depiction of PALETTE processes and outcomes, PALETTE., Pg 4

[47] (2007), WP6 summary formative report 4: challenges, connections and awareness, Oct 2007.

[48] (2008), On formative evaluation as a lever for change, Feb 2008.

[49] Op. Cit. Pgs. 5 - 6.

threatens the wider activity or actors. Seen in terms of structure and action, evaluation could dysfunction if the relationship between a part of the structure and a part of the context or the actions being carried out does not respond to expectations. One would expect this to lead to increased complexity requiring additional measures to simplify things again.

To attempt to understand such a way of seeing things let's take two examples. Let's start with the work of the Steering Committee at the beginning of the project when it had to select and adopt an audio-conference system. For a brief description of this moment of evaluation consult the annexes. The organisation of the evaluation in this case could be loosely divided up into three parts: the identification of the problem and adopting measures to respond to it; the identification and testing of potential solutions; feedback and decision-making.

The procedure adopted is informal, both in terms of how the overall evaluation is organised and how it is monitored and feedback is provided. This leaves considerable freedom for those wishing to use their own practices. At the same time, although there is a clear decision to act and the results are briefly presented and noted in minutes, giving an impression of explicitness, the nature of the evaluation is highly implicit in that no discussion takes place about how the evaluation is to be carried out. The level of embedment is not so clear. There is an implicit assumption that a common audio conferencing system will be found, at least for Steering Committee meetings, but individuals and groups feel free to experiment with and adopt other systems.

Now let's look at each of the three parts of the evaluation in terms of their action and interaction with the context and see where there is potential dysfunction and increased complexity, or, on the contrary, the structure of the evaluation serves to diminish complexity and, as a consequence, improves understanding.

1. **Problem identification and the design of measures**

   This activity focuses the attention (and the energy) of Steering Committee members on a specific problem and should engage them in deciding on measures to take. In evaluation terms this would correspond to the design of the evaluation. In practice, responsibility was shifted to a small number of partners who spontaneously volunteered to "look into" the question and "find" a solution. No 'design' was made for the measures to be taken. Individual actors implicitly referred to the practices that were already familiar to them. This delegation of responsibilities meant that most of the SC members disengaged from the problem and finding its solution even though they were directly concerned with the need for a solution. Neither this disengagement, nor the lack of a shared design were necessarily a problem in terms of ultimately finding a satisfactory solution, providing there was a satisfactory follow-up on the evaluation.

2. **Identify and test solutions**

   This stage, in evaluation terms, corresponds to the heart of the evaluation process where information is gathered and hypothesis tested. It should also channel energy into finding a solution. It was not, however, explicitly structured, leaving those who had accepted the task to decide how they carried it out. Thus giving free reign to a potential variety of professional practices, but with the risk of possible misunderstandings between those practices. In addition, a number of other people and groups also tried out solutions without necessarily consulting their colleagues. As a result there was a possible non-alignment of the effort of the various actors to find a solution. This in itself need not have led to a dysfunction because it could have produced a richer set of knowledge on which to decide.

3. **Feedback and decision-making**

   The process of identifying and testing solutions was not documented except in a brief note in the SC minutes. As a result, feedback was informally provided during an SC meeting or possibly informally in private conversations. To what extent did this relative absence of reification hinder the process? Of course, a dysfunction at this stage could also be due to the absence of a suitable technical solution on the market.

This whole procedure and the related evaluative activities correspond to a certain culture of management that might be called 'laisser-faire'. As mentioned above, it could well have produced satisfactory results in this particular case that wasn't necessarily very complex. However, there were signs of some dysfunction, as the process of selecting a solution took a long time and a number of trials of audio-conference solutions in real-life situations led to communication problems for the groups trying to use them.

The second example is an evaluation carried out by WP6 during the Lyon Plenary Meeting at the end of 2007. Its purpose was to improve understanding amongst Steering Committee members of formative evaluation in the PALETTE project and its relationship to the PALETTE Participative Design Methodology (PDM). A brief description of this moment of evaluation can be found in the annexes. This evaluation was largely formal in that design of the evaluation and the procedures for collecting data were clearly laid out prior to the evaluation even if the latter did leave room for the unexpected. It is interesting to note that the analysis and formulation of the results, however, were not dictated by an explicit formal procedure. The intention of the evaluation was explicit: apart from seeking to develop formative evaluation it might also be said, more implicitly, to encourage empowerment through the development of reflexive practice on the part of participants. The evaluation was partly embedded in project activities through its interaction with the work of the Steering Committee and the project management. Finally the procedure was consultative and, to a limited extent, participative, involving participants in validating prior findings about formative evaluation and revising and debating those findings with them during the interviews and subsequently during the presentation of the results to the Steering Committee.

Seen in terms of its relationship to the actors and the project environment, this evaluation had a number of "phases" to it.

1. **Preparatory documents**

   Steering Committee (SC) members were sent a one-page preparatory document containing a brief summary of earlier feedback from them about the uses of formative evaluation[50]. The document also contained a list of five 'prompts' centred on personal experience of evaluation to "*prompt thoughts on the uses of evaluation*"[51] prior to the interview. This document was to pave the way for improved feedback and also served to focus the attention of SC members on the question of formative evaluation. Providing a summary of earlier feedback potentially enabled SC members to compare present perceptions with their earlier opinions. The designers of the evaluation were aware that providing this summary might unnecessarily limit the input on evaluation by 'colouring' people's answers, but decided it was worth the risk. On a longer-term basis, providing these preparatory documents could be seen to prepare the presentation of the subsequent results to SC members by raising awareness and stimulating debate about formative evaluation.

   The success of such a strategy of involvement, however, depended partly on the extent to which interviewees took the time to read the material sent to them and reflect on it. A number of interviewees came to the interview with annotated copies of the prompts. Others admitted not having read what was sent to them and several people misunderstood the meaning and purpose of the documents.

2. **Interviews**

   The interviews themselves were carried out individually and took place during the Plenary Meeting, that is to say, in a context of intense exchange and reflection about past work and the work of the remaining year of the project. Interviews lasted between twenty and thirty minutes. Although the interviews were structured around a series of five 'prompts', some flexibility in handling them[52] meant that participants were able to elaborate and expand on their ideas, should they wish to. Many interviewees welcomed this opportunity to dialogue about the subject and about the project. The flexibility of the interviews permitted the emergence of a key subject that had not been foreseen by the evaluators: the relationship between evaluation and decision-making.

3. **Analysis**

   In the analysis of the interviews, they were rendered anonymous as part of an explicit 'contract' to enable people to speak freely. In the report, this helped shift focus away from personal opinions and possible conflicts or tensions. Anonymity had a cost, however. *"…. Important information contained in the link between the statements and the person's role in the project were passed over in silence.*"[53] Seen from another point of view, anonymity also

---

[50] The document said: "*Below are some notes on feedback about formative work from the coordination of the project and other members of the steering group.*"

[51] Both quotes come from the unpublished document sent to SC members prior to the interviews.

[52] The approach was akin to so-called "non-directive techniques" in which the interviewer echoes back something the person has said with a view to clarifying or extending the ideas.

[53] McCluskey A., On formative evaluation as a lever for change, PALETTE WP6, June 2008.

implied that individual members of the SC partly lost any ownership they might have had of what they said.

A major task of the analysis consisted of categorising and re-organising the material in such a way that the complexity of the multiple opinions and perceptions was diminished without loosing any essential ideas and, at the same time, still reflecting the diversity of perspectives present in PALETTE. A series of recommendations were made, many of which came from interviewees themselves.

4. **Presentation and discussion in SC**

The report was deliberately kept short: five and a half pages. It was sent to Steering Committee members and briefly presented during a Steering Committee meeting two months after the evaluation had taken place. The main reason for providing feedback to the SC was to inform decision-making concerning the project and improve understanding of formative evaluation. In addition, as mentioned above, the fact that answers were made anonymous resulted in a loss of ownership on the part of interviewees. Discussion of the results within the SC gave a potential opportunity for some re-appropriation of the various discourses.

The question at the end of our analysis of this moment of evaluation remains the same as one of those asked of interviewees: that of the impact of the evaluation. To what extent did this structure (and design) of evaluation have an impact on the use of knowledge developed during the evaluation to improve the understanding of participants, their ways of working and the management of the project? And what might have been changed in the structure of the evaluation to improve its use? This is a very frustrating and possibly misguided avenue of exploration. Why?

First of all, there is the assumed relationship of causality between 'design' and use. If, for example it had been possible to hold a half-day workshop with Steering Committee members on the outcomes of the evaluation, would that necessarily have improved the use of the results? It would certainly have signified a will to take the evaluation seriously, but it is less sure that 'use' would necessarily have been improved.

The second difficulty lies in the notion of improvement. Within the perceptions of Steering Committee members, as expressed individually and in private, quite differing and even divergent opinions and frameworks were expressed. Improvement for one person would have entailed increased intervention on the part of evaluators. For another person, evaluation should limit itself to a "passive" role of observation. These multiple perspectives about the role and functioning of evaluation, despite earlier efforts to reach a common understanding on formative evaluation, imply that 'improvement' is a difficult standard with which to measure change. One could posit an external set of criteria to measure change but ultimately the appropriateness of that change depends on the perceptions of those taking part, not on external criteria.

Finally there is the question of 'impact' itself. The dictionary tells us that the word 'impact' has its origins in the concept of 'impinging' and its use dates from the seventeenth century, the time of Newton. Our conception of the world is still largely influenced by Newtonian physics via such principles as action and reaction or cause and effect. We now know that Newtonian physics is only a local approximation and that complex phenomena do not comply with such laws. But for everyday purposes we continue to think the world in Newtonian terms. And our expectations, including those related to evaluation, are coloured by the thought that every action produces a reaction and that every cause has a discernable effect. There is a need to rethink the question of the relationship between design and use in evaluation in terms that are compatible with the complexity being handled.

Reification

To what extent are the processes or the outputs of these moments of evaluation reified? Do participants write down what they do, or record their proceedings in some other way? Or would doing so unnecessarily hamper the process or make it too cumbersome? Is thinking about the way things are done seen as a welcome learning experience or is it avoided as a nuisance or a possible hindrance to activities? Looking at these questions from the perspective of the body of theory about communities of practice, both learning and identity are seen to be reinforced through reification[54]. To what extent does or would reification contribute to a better understanding of moments of evaluation

---

[54] Wenger E., Communities of practice: learning, meaning, and identity, Cambridge University Press, Cambridge, 1999

and their improvement both in terms of professional practices of a given community and as activities within the project in which they are integrated?

Reification is not necessarily bound to explicitness or formality. A process might be explicit, i.e. clearly stated, but what actually happens may remain unrecorded and unchallenged. The same goes for formality. The process may be dictated by a set of rules or procedures yet the process itself may not be written down or otherwise recorded or questioned. Formality (or possibly familiarity) might lead rather to the non-reification of processes because they are supposed to follow pre-determined and frequently used procedures. We will talk more about this below when we discussed the relationship between espoused reality and what might be called 'actual' reality.

Let us look at the example of the organisation of the Scientific Advisory Board (SAB). A brief description of this moment of evaluation can be found in the annexes. Although organising meetings of the SAB is an explicit activity with the intention of gathering feedback from its members on specific issues, the process is not formalised, that is to say there is no procedure laid down for how meetings are convened, subjects chosen or discussions organised. These aspects are left to 'common sense' and are likely to include practices familiar to some or all of those participating. They are also probably dictated by contingencies of the moment. For example, the absence of one external member led to a written commentary being submitted but this practice was not repeated in the subsequent meeting despite the fact that it provided useful feedback in a structured, easily sharable form.

Another question about the organisation of meetings concerns the processes by which decisions are made about subjects to be discussed. There is no formal procedure and no post hoc description of the process. It would seem that there is a tacit delegation of this responsibility to one or several people. Presumably this situation is judged sufficient because no mention was made of problems due to the choice of subjects.

Minutes of the two SAB meetings are available on the project document server (the BSCW). The minutes describe succinctly what happened, but according to one key participant they do not reflect the richness of the discussions. The minutes give rather the impression that PALETTE partners spent more time explaining their position than listening to what external members of the SAB provided as feedback and suggestions. This misleading impression does not contribute to strengthening the image of the project or the work of the SAB. Nor does it reflect the good will of some participants to try to have more extensive notes produced. It is amusing to note that, in a project actively involving discussions with CoPs about the need to reify their practices, no explicit thought is given to why and how the work of the SAB should be reified. It is possible that for the correct functioning of the project, extensive notes on ideas generated at SAB meetings were not necessary. SAB meetings were often organised in conjunction with SC meetings and many SC members were present, so individual memories of discussion might have been sufficient to aliment decisions during SC meetings. On a longer-term basis, however, it is less sure that personal memories were sufficient when it was necessary to refer back to relevant SAB discussions that became pertinent much later. For example, in written input to the Feb 2007, one SAB member wrote: *"... possible dangers will arise if the project becomes too fixated on meeting goals and too bureaucratic, and fails to be open to new insights and changing directions.*"[55] This assessment echoes concerns two years later when some interviewed SC members complained that the project work was too dictated by deliverables to the Commission[56].

It would seem that the SAB functioned well, produced interesting feedback and contributed to preparations of both the official review procedures and the revisions of the implementation plan (IP). From what we've seen above however, the articulation with the project and in particular the steering committee and the project coordination worked less efficiently. It is not so much that this produced increased complexity (as mentioned in other examples above) but rather than it contributed to entropy in that valuable information and knowledge might have been lost or forgotten because it was not kept in a more sustainable and accessible form.

Espoused and actual reality

With 'espoused' reality there is an openly shared discourse about the way reality is said to be, whereas 'actual' reality refers to what can be observed as taking place. The distinction becomes interesting when the two realities do not correspond, when the discourse about reality does not coincide with what one observes and, furthermore, efforts are deployed to protect and maintain the

---

[55] Written input from Gord McCalla.

[56] While gathering material for: McCluskey A., On formative evaluation as a lever for change, PALETTE WP6, June 2008

espoused reality. This situation becomes critical when the gap between espoused and 'actual' reality widens and when decisions are based only on espoused reality.

Note, en passant, that there are a number of underlying questions that have to be addressed about these concepts, although we won't go into detail here. First, to what extent is espoused reality not really a shared discourse but rather a convenient concept to which people pay lip service? Second, are there not various levels of discourse instead of a unique espoused reality? Third, given that 'actual' reality depends on observation and the results of observation reflect personal and collective perceptions that are influenced by a number of external factors, to what extent is there a unique 'observable' reality to oppose its espoused cousin? There are many informal and often implicit shared expectations and 'frames of reference' that dictate what is expected and how 'reality' is evaluated. Finally, there are indications within PALETTE, given its multinational consortium, that the status of espoused reality, especially in the form of the written word, as synonymous or not with actual reality may vary from culture to culture. For some people, the written reports and project deliverables are a faithful indication of the actual workings of the project, others are more sceptical.

Returning to the distinction between espoused and actual reality, let's pursue with three examples from the PALETTE project. 'Espoused' might refer to an explicit, formal discourse about reality, for example a theory or a work plan or the description of an activity on which people openly agree and which is officially recognised. The Description of Work (DOW) for the PALETTE project is an example of such an espoused reality. To take into account the fact that the reality might change over time and the progress of the project might require readjustment the European Commission makes provision for periodic official modifications through the review and revised implementation plan processes. These revised work-plans remain an espoused reality, replacing the former version. They are also contractual in nature, requiring compliance on the part of project partners with respect to the funders and as such, espoused reality should not differ from what one could call the 'actual' reality, that is to say, what in practice happens.

The formal Commission review process, as a form of summative evaluation, sets out to test the alignment between the espoused project reality as expressed in the work plan and the observed reality based on project deliverables, other project documents and the discourse of project participants. Compliance then is largely measured in terms of deliverables: has what was promised been provided? In terms of discourse and observed reality, the question arises as to the extent to which deliverables reflect the espoused discourse or are an indication of the observable reality of the project. Given the considerable pressure present in the preparation of deliverables to ensure compliance, to what extent might they not tend to be a further statement of the espoused discourse, distancing the external evaluators from the 'reality' of the project? This subject is understandably fraught with difficulty because, as long as projects are judged on compliance with promised work and funding depends on it, it is risky for project partners to evoke a possible divergence between espoused and actual reality.

The espoused reality might also be informal and tacit, for example when a community or a set of project partners share a vision of the way things are without that vision necessarily being expressed in writing or being the subject of a formal negotiation. André Gorz[57] wrote that one of the most important contributions of utopias to society is that they indicate a shared intention and a direction in which to go. Although they may not be classified as utopian, there are likely to be a number of belief sets behind large projects like PALETTE that guide and give direction to the project. The Participatory Design Methodology (PDM), for example, is the subject of much deliberation and a number of lengthy deliverables, however, one might considered that it constitutes a tacit espoused reality in that project partners tacitly adhere apparently unquestioningly[58] to the idea that a participatory approach to the design of tools and services involving programmers, pedagogues and users (CoP members) is a more efficient way of developing and adopting those services. Whereas the actual discourse of project

---

[57] André Gorz, Capitalisme, socialisme, écologie, Galilée, 1991

[58] In response to this statement, a member of the Steering Committee proposed an interesting alternative vision. Note that his vision does not contradict what is written above: "*This example is easily understandable with the ANT concepts of 'alignment' and 'black box'. At the beginning of PALETTE, a group of researchers developed the model of PDM while other groups proposed other schemas and frameworks. The discussions led to the tacit adoption of the current MOT schemas. The others tacitly accepted because the discussions would have been too long and there were deadlines to meet. So the 'alignment' of the others on the proposal was not democratic and occurred 'by default'. After this episode, the schema became a 'black box' i.e. an object that is discussable but that is not discussed because it would have been too time-consuming and asked too many efforts for having an agreement.*" This pressure of time and deliverables is often used as a major argument against challenging underlying assumptions and choices.

members when questioned about the PDM in the Lyon evaluation on formative evaluation and the PDM, indicated a wide diversity of perceptions of the method and of its effectiveness.

On a different level, the espoused reality of PALETTE includes the idea of a considerable degree of autonomy of the WPs in running their work[59]. One might expect this autonomy to include choices concerning evaluation, but it was not always reflected in participants' expectations with respect to evaluation. Output from the Lyon interviews indicated considerable ambivalence about the subject. A number of Steering Committee members expected the evaluation work package to intervene in the other work packages providing means for evaluation and even executing that evaluation, while others were not disturbed by the non-intervention of the evaluation specialists in their work. This ambivalence was to be found within the evaluation work package as well. This situation might be partially explained by the coexistence of two differing perceptions of evaluation: one, explicit, in which evaluation is seen to be the province of professional evaluators; and the other, implicit, where evaluation is part of the practice of varying professional groups and not just evaluators. One of the central results of the current report is to make the cohabitation of those two perspectives explicit.

As we have seen above, many of the moments of evaluation taking place in PALETTE are heavily embedded and functional. That is to say they provide essential knowledge needed for the particular activity they are embedded in, and as mentioned elsewhere here, the more integrated the work the more likely it is to go hand in hand with limits on what can be called into question. As such these evaluative moments are less likely to uncover discrepancies between the espoused discourse and the actual activities of the project. To illustrate the inability of embedded evaluation to uncover discrepancies between espoused and actual reality would require showing that 1) a discrepancy existed and 2) the form and nature of the evaluation prevented unearthing and exploring that discrepancy. As a result, any example cited here could only be seen as a hypothesis for various reasons, one of which is the inherent risk in pointing to discrepancies between espoused and actual reality when the absence of discrepancy is seen as a key criteria for compliance with funders conditions. Secondly, for cultural and identity reasons, written reports might be taken as an indication of actual reality when they may well be a form of espoused reality that does not necessarily fully reflect actual reality.

---

[59] See the official Description of Work (DOW).

4. Possible ways forward

The following paragraphs raise a number of the outstanding questions and point to possible ways forward that are subsequently summarised in the final section in the form of fifteen points "to take to work".

Four perspectives

The reader may have noticed that a number of perspectives are present in this report. It is important to evoke them here in an attempt to avoid potential misunderstandings due to their cohabitation. Four perspectives play a predominant role. The first might be called 'philosophical' in that it reflects on the concept of evaluation and seeks to extend and enrich that notion. It doing so, it asks questions about words and meanings. The second might be called 'academic' in that it looks at those extended forms of evaluation and seeks to construct understanding on the basis of structured observations. In line with what might be called a 'territorial approach' to knowledge, as reflected in many academic structures, it seeks to clarify which body of knowledge, which methodologies and which perspectives it is approaching the problem from. It draws its legitimacy from this anchorage in a given body of academic practice. The third perspective that might be called 'change management' looks at both individual and collective strategies of change. It looks at organisations and policies and trends. Theory may be a source of legitimacy for change management, but what change management seeks above all are images and metaphors with which to make sense of the world, to fire the imagination and to build change. The territoriality of the academic approach might be seen as a barrier from the perspective of innovation. The philosophical questioning might also be seen as a barrier to action. Finally the fourth perspective is that of the actors carrying out the activities that we have called 'extended evaluative practices.' Primarily they have processes to carry out and goals and targets to reach. Using what we have labelled 'extended evaluative practices' is a means for them to get their work done. They might or might not reflect about the nature of their practice. They might or might not try to get a better understanding of what they are doing. They might or might not be concerned about learning and improving their practices. The advocate of the change management perspective might argue that reflection, understanding and improvement are essential to doing things better. The academic, coming from the perspective of 'social practice', might argue that looking at such activities in terms of practice and communities of practice provides a coherent picture of what is happening, reaching the conclusion, for example, that the reification of that practice is a good strategy for learning and improvement.

As mentioned above, we underline these differing perspectives because the evaluation reached from one perspective and the recommendations that might be made on the basis of that might not be appropriate for those who work according to a different perspective and might even lead to misunderstandings. Proposing the reification of processes as a strategy for improvement is a good example. From a theoretical perspective such a strategy seems sound. In practice, it is not so sure that reification is a solution that can be generalised especially when it comes to tacit and informal processes. Being aware of the existence of these different perspectives is a possible first step, in a trans-disciplinary context, towards being able to transform such suggestions into suitable propositions from one's own perspective and practices or to understand why one's theories wouldn't work in somebody else's practice. Such a 'mutual understanding' activity, however, can only work if it is recognised as necessary and legitimate by those in control, if it is taken seriously by all those involved and if time is allotted for it to take place.

Cohabitation between experts and the lay

From the perspective of the community of evaluators, the extension of the notion of evaluation to include what we have called evaluative moments may be seen as a challenge, if not a threat. It is not meant as such. We have discussed above the question of the legitimacy of the lay perspective and the inevitable tension between experts and the lay. Writings about communities of practice[60] underline the importance of the feeling of identity that goes with belonging to a community and that is built around learning and sharing. In our observations it was clear that those people who use moments of

---

[60] For example: Wenger E., Communities of practice: learning, meaning, and identity, Cambridge University Press, Cambridge, 1999

evaluation in their work but who are not professional evaluators do not see themselves as evaluators. Their professional community is elsewhere: developers, pedagogues, group animators, …

There is however a challenge here: using the opening of evaluation to include moments of evaluation as an occasion to pursue topics like the usability of evaluation (see below) but also to consider the cohabitation of expert evaluative practices and those of the lay using certain evaluative practices. Considering evaluation as a series of practices may help in making explicit the limits and possibilities of both 'externalised' evaluation as well as embedded evaluative practices that occur as part of sets of practices associated with the development or processes that occur within project activity. In the initial elaboration and description of projects (including those for funding purposes), both types of evaluation should be depicted as part of an 'evaluative or reflexive culture' with a view to encouraging them during the project.

Evaluation as a means to handle complexity

We have suggested in this report that effective evaluation contributes to diminishing perceived complexity, if only temporarily, and as a result can facilitate decision-making. Saunders and others have talked of evaluation as depicting 'temporary stabilities'. Actor Network Theory speaks of the emergence of 'actors' that serve to simplify the system. However, as mentioned above, most people continue to understand the world in terms of Newtonian physics (even if they might not call it so). The notions of complexity and emergence have still to penetrate many decision-making processes, which, as a result, remain inadequate because they can't satisfactorily handle complexity. Explicitly linking evaluation and complexity with suitable images and metaphors, may help to both popularise ideas about complexity and help understand the evaluative role in many practices.

Focussing on usability as a lever for improvement

The concept of 'usability' is an attempt, from the academic perspective, to subsume identifiable practical aspects of extended evaluative practices that might be assessed and improved. From a semantic perspective, the dictionary says of usability: *able or fit to be used*. Other words spring to mind: suitability, efficiency, impact, appropriateness, value for money,…, each with its own slant on the question. The word 'usability' is interesting in that it necessarily points to the way something is used and as such addresses the relationship between the 'design' of the activity (even if it is entirely tacit and informal) and the use to which it is put seen in the light of its role and place in a wider context or process. The articulation between the evaluative moment and the wider process, as described elsewhere here, is a key facet of usability. The concept of 'usability' is also interesting in that it can be of use from a change management perspective. It is also, to some extent, close to the more practical preoccupations of those carrying out these 'evaluative' activities. So why such interest in 'usability'? Because innovation, improvement, development, adaptability and learning are key concepts in current political, industrial, ethical and pedagogical agendas. And the concept of usability, in the case of extended evaluative practices, might be a way of approaching the question of changes in practices and learning from those changes that can be assimilated more readily by the 'practicians' themselves and their professional communities. More work needs to be done identifying aspects of usability in terms that are meaningful to practicians.

A close examination of the examples provided here indicates, however, that there are limits to improvements that can be made via the concept of usability. Why is this? Many of the aspects of the moments of evaluation that might be singled out for possible improvement can be traced back to factors that are rooted in the wider context (structural considerations, contractual constraints, budgetary limits, differences of professional culture, the personality of key actors, differences in regional or national culture, …). As we have mentioned elsewhere here, the embedded nature of many moments of evaluation implies that such meta-level questions cannot be addressed in the context of those moments. What's more, many of these 'larger' questions often can't even be addressed in the wider framework of the project as a whole.

The need for go-betweens

This report has illustrated the extent to which activities related to evaluation play a key role in many professional practices. It has also underlined the diversity and the complexity of those activities and their relationship to other activities within the project and beyond it. At the same time, the level of awareness of these activities as 'evaluative' is often very low both on the part of those people carrying them out but also on the part of professional evaluators. It is not always useful to point out to people that what they are doing might be quite different from what they think. The examples given here make a case for increasing awareness of the evaluative nature of some practices as a possible path to

improved 'usability'. The central question remains how best to enhance that awareness. Thinking about social practice and the framework of communities of practice point to the innovative impact on practices of those who move at the periphery of such communities, acting as 'go-betweens' between differing communities and different practices and perspectives. Note that these go-betweens are not external experts, but rather people who have one foot in and one foot out a given community. The position of such people is not always comfortable because communities tend to want to bring those at the periphery fully within their logic. Focussing more attention on the role of go-between might be a viable strategy for organically raising awareness about new practices and new ways of seeing practices.

Where are the jesters for espoused reality?

In the times when kings and queens reigned, each court had its jester, someone who was permitted to stand outside conventions and, thanks to the mantle of humour, was able to poke fun at the ways things were and point to otherwise unnameable possibilities. It was a delicate task and jesters often lost their heads in the process, but the risk was worth it. Where are the jesters to make us cry with laughter over the ways of international research projects? In talking of espoused reality above and its relation to 'actual reality', we evoked the difficulty, if not the dangers, of uncovering growing gaps between the two. We also mentioned the difficulty, within the framework of a given evaluative activity, of questioning the underlying assumptions on which those activities were built. There are clearly places and times when challenging assumptions and ways of working is not welcome. It might be fruitful to explore 'places' and roles in which transgression would be possible, if not encouraged, and where that transgression could be innovative and constructive.

5. Takeaways: Key points to take to work with you

The following fifteen points are given as possible 'instruments' that can be used to question ways of working, to stimulate reflection about the place and role of evaluation in project work and to encourage innovation and improvement in evaluative practices in complex R&D projects.

1. Evaluation helps handle **complexity** by providing provisional stabilities that assist decision-making processes and can serve to combat entropy by enabling the emergence of a simpler order from complexity, albeit temporarily.

2. Understanding and improving evaluation in complex projects requires extending the notion of evaluation to include **moments of evaluation**, that is to say, embedded evaluative processes not necessarily carried out by evaluation experts.

3. Many people are unaware that evaluation is an activity that **involves and concerns all participants** in a complex project like Palette and not just the experts of evaluation. Evaluative moments are invariably to be found amongst their professional practices.

4. Moments of evaluation can usefully be described in terms of a number of **phases** including: the negotiation with the actors of the work to be done, the process design and the creation of tools, the collection or generation of data, the analysis and/or reorganisation of that data and the integration of this (new) knowledge in a process.

5. Making the integration of evaluation outcomes part of the evaluation moment links evaluation more solidly to **decision-making** and raises questions about the relationship to decision-making in conventional notions of evaluation.

6. The understanding of this extended notion of evaluation can be enhanced by looking at evaluation as **a series of practices** developed within one or more communities of practice.

7. As evaluative practices may differ from one community to another, working in a trans-disciplinary project like Palette requires that attention be paid to **differences of perspective** resulting from evaluation as potential sources of misunderstanding but also of learning.

8. These differences in perspective need to be **'surfaced'** in order to collectively understand, appreciate and manage them in a productive way. Surfacing is different to the reification process we identify below in that it is more about communication than capture.

9. The difference between **experts and the lay** in evaluation can be understood from a social practice perspective in that evaluation experts belong to communities of practice centred on evaluation whereas lay users of evaluation belong to communities not centred on evaluation but whose practices contain some activities related to evaluation.

10. In a **participative** project centred on knowledge like Palette the relation between experts and laypeople is necessarily problematic because of the imbalance of power due to differences in the perceived legitimacy of their respective knowledge and expertise.

11. The practice perspective throws new light on the way knowledge is given form and shape and is made to last (**reification**) as a process of learning. It raises the question of the extent to which evaluative processes and their outcomes are recorded and made available and whether this might contribute to improved efficiency or impact.

12. Understanding evaluation from a practice perspective suggests ways in which innovative practices can be adopted thanks to the efforts of those people who gravitate at the limits of such communities and act as **go-betweens or knowledge brokers**, bringing new ideas with them. These activities need to be encouraged and enhanced if we seek to innovate in ways of working.

13. The concept of **usability**, that addresses the relationship between the 'design' of the evaluative activity (even if it is entirely tacit and informal) and the use to which it is put seen in the light of its role and place in a wider context or process, is a useful lever for improving evaluation and evaluative moments.

14. Usability has its **limits**, however, in particular the extent to which evaluation is able to question the framework in which it is taking place.

15. To enable the critical evaluation of fundamental questions, assumptions and frameworks, thought should be given to creating 'places' and roles in which **transgression** (asking questions that are generally not allowed) would be possible, if not encouraged, such that transgression could be innovative and constructive.

6. Annexes

Examples of moments of evaluation

The following table indicates where examples of moments of evaluation have been used in Sections 2 and 3 of this report.

	Section	Examples
2	Embedded – detached	WP1 – Authentification of scenarios
2	Explicit – implicit	
2	Formal - informal	WP0 – Ways of working & communication systems
		WP0 – Deliverable evaluation process
		WP0 – Review and IP process
		WP1 – The Trials
2	Expert – Layperson	WP1 – Authentification of scenarios
2	Hierarchical - Participative	WP1 – Authentification of scenarios
2	Complexity and provisional stabilities	WP6 – Lyon evaluation
2	Entropy	WP0 – SAB
2	Communities of practice	
3	Multifaceted nature of evaluation	WP0 – Ways of working & communication systems
		WP0 – Deliverable evaluation process
		WP0 – SAB
		WP1 – The Trials
		WP6 – Lyon evaluation
3	Structure and action	WP0 – Ways of working & communication systems
		WP6 – Lyon evaluation
		WP4 – Internal testing of scenarios
3	Reification	WP0 – SAB
3	Espoused and 'actual'	WP6 – Lyon evaluation

Deliverable evaluation process (WP0)

A project like Palette produces a great many official deliverables. Such deliverables are seen first and foremost as produced for the European Commission as part of the project management process in the framework of EU funding[61]. Procedures for the writing and review of these deliverables were discussed by the Steering Committee (SC) during their meeting in July 2006 in the middle of an intense period of activity preparing the first major set of deliverables. The minutes of the SC meeting indicate that guidelines were to be fixed for the review of deliverables.

Each official deliverable has at least two evaluators. Most evaluation is carried out by internal evaluators, although key deliverables can also be submitted to external evaluators. The SC meeting in March 2007 discussed procedures for choosing reviewers but the task was put off till after the first review. The process was further discussed at the July 2007 but no formal decision was taken.

When the deliverable is complete, it is sent to the reviewers for comment. Approaches to reviewing vary from reviewer to reviewer. Most make comments on a separate document, making reference to relevant place in the original. Others make changes in the original document using the function in Word or PDF to trace changes. The reviewers' remarks are posted on the BSCW and, if necessary, the deliverable is revised. It is then finally submitted to SC for approval (along with the reviews) before being sent to the Commission.

---

[61] This status of deliverables and the onus put on producing them rather than carrying out the project was pointed out by a number of participants as problematic during an evaluation carried out in Lyon in December 2007.

Ways of working and communication tools (WP0)

Probably some of the least formal moments of evaluation that took place in the Steering Committee (SC) concerned shaping ways of working and developing forms of communication. At the beginning of the project, a number of aspects needed defining and refining such as the use of the web site and internal communication using mailing lists, audio & video conferencing and the communications platform (BSCW). Let's take the example of the Palette website. The website went public in May 2006. The site was then discussed on several occasions by the SC in June and July 2006. SC members were asked to send comments and suggestions. No formal structure or set of questions were used to solicit reactions. A formal deliverable (D.DIST.01) was submitted to the Commission mid-July that presented the web site in paper format with a series of screen shots.

Another example of this « informal » evaluation of communication tools was the testing and trialling of conferencing systems. A number of solutions were tried out in real-life conditions by the SC or by individual work packages. Mention is also made of testing the EFPL conferencing tools and brief feedback about these tests is noted in the minutes of the October 2006 SC meetings. The minutes of September 2006 meeting of the SC mention an FAQ about such tools. However the PALETTE FAQ published at the end of October 2006 makes no mention of video or audio conferencing. And the planned[62] « PAL-Internal_Tools » document is not on the official project document store (BSCW).

Scientific Advisory Board (WP0)

The Palette Scientific Advisory Board (SAB) is composed of a limited number of external experts who meet once a year with key players from the Palette project. In the minutes[63] of the April 2008 meeting of the SAB, the goal of the work was said to be: "*to collect feedback from experts in order to take strategic decisions and prepare the next review meeting.*"

SAB meetings are structured around a set of key issues that are decided on by the Scientific Coordination. Oral (and occasionally written) feedback from SAB members takes the form of requests for clarification, indications of possible shortcomings, discussion of potential risks, presentations of sources of information and synergy, suggestions for future developments, … Outputs of the work of the SAB are internal to the project and the Commission evaluators and are available in form of "minutes" as well as the presentations provided.

Reading the minutes available on the BSCW[64], it seems apparent that the SAB also served as a forum for discussion amongst Palette partners about key issues. The minutes give the impression that this activity dominated discussions, as input from SAB members appears comparatively limited. Questioned about this, one key player in the Steering Committee said that the written minutes did not reflect the richness of the discussions.

The review and IP process (WP0)

Two key moments[65] in three-year project management process of PALETTE were the official reviews by a team of experts appointed by the Commission. These reviews were associated with a revision of the Implementation Plan (IP) that was subsequently negotiated with the Commission. This review process and the modification of the Implementation Plan played a multiple evaluative role. The review is a summative process in which delivered work is approved by a group of experts on behalf of the Commission and the pursuit of the project is decided on. The elaboration of the revised IP serves the longer term steering of the project. And, finally, the whole process plays a formative role as participants learn from what has been done and from the comments of the reviewers so as to improve ways of working and communicating.

Taking as example the first review and the preparation of IP2, the SC meeting in January 2007 dealt at length with both the report of the first year of Palette drawn up by project partners as well as propositions for IP2, building, amongst other things, on the evaluation report from WP6 presented at the December meeting of the SC. In her overview of the first year of the project as well as in her presentation foreseen IP2 activities, the project Scientific Coordinator highlighted a number of issues that needed to be addressed particularly in reformulating the Implementation Plan. It was suggested that work on the new IP should proceed bottom-up: from tasks to Work Packages (WPs).

---

[62] Mentioned in PAL-SC-05-V2.doc

[63] SAB-Minutes.doc

[64] February 2007, November 2007.

[65] Early 2007 and early 2008.

The process of preparing the review included an initial self-assessment of their work by WP leaders. Members of the SC were reminded that their internal assessment for the first review meeting should refer to the twenty-four impact and results indicators drawn up in the DoW. The impact indicators are seen in terms of the audiences addressed (individuals, CoPs, organisations and society) and the results indicators are organised in five categories (producing, using and sharing information; reification of knowledge about practices; supporting collaborative learning; interoperability of tools; and usability, utility and acceptability of tools). Note that it was suggested that this list of indicators was too complex and be shortened in IP2.

Three subsequent meetings of the SC (in February and twice in March) were largely dedicated to working on the review and IP2. Input from the SAB meeting that took place in February 2007 was also incorporated into the process.

Authentification of scenarios (WP1)

The Participative Design Methodology (PDM) developed in PALETTE entails three main stages. The first, "analysis phase" concerns analysing the situation including existing tools and activities in CoPs. The second, that could be called "participatory design for use" involves the design of potential new scenarios and developing related prototypes. The third and final phase, "participative design in use", centres on trialling scenarios and services and modifying these according to the results.

In the second of the three phases of the PDM, the participatory design for use, a more complex authentication process was employed to design new activity scenarios for CoPs. Not only were the scenarios authenticated in terms of possible usage by CoPs, but they were also authenticated in terms of usability and technical feasibility. The design of scenarios, which is described in Fig 15 of "Refining and Instrumentation of the Participatory Design Methodology", involved several stages and began with the conception of scenarios taking into account the description of CoP activities developed in the analysis phase of the PDM. The design of scenarios was carried out by teams of CoP delegates, CoP mediators, programmers and an external "authenticator"[66]. The "authenticator" acted as a "critical friend" who could provide an outside perspective. Subsequently, the scenarios were authenticated by CoP delegates or focus groups. Authentication was conducted using a questionnaire developed by the authenticator and the CoP mediators on the basis of the generic indicators provided in D.EVA.02. The authenticator wrote a synthesis of the data thus collected and the results of the work were published in D.PAR.03.

The scenarios that had been authenticated by the CoPs were then further authenticated for their technical feasibility according to criteria listed in D.PAR.03. The results of this feasibility study were published in the same document.

The Trials (WP1 & 5)

The so-called Trials are probably the most complex evaluation moments in the PALETTE project. Starting from scenarios developed by teams composed of CoP mediators, developers and representatives of CoP in the context of the work package dedicated to the implementation of PALETTE services and scenarios (WP6), based on both the needs and practices of the CoPs and tools and services being developed in PALETTE, the subject of the Trials is the use of available tools in the context of the scenarios by CoPs in conjunction with an on-going dialogue between users (CoPs) and developers (the 'technical' work packages, WPs 2, 3 & 4). The trials themselves are a period of observation of that use followed by analysis and dissemination. A group of experts from WP1 elaborated a protocol for the observation of use, including such conceptual axes as instrumentation, instrumentalisation and mediation. The observation and analysis are organised and carried out by mediators, that is to say people who are either experts in the working of a CoP or someone with a good overview of a tool or set of tools. Each mediator or group of mediators organises the observation as an independent research project following a protocol inspired by the general protocol mentioned above but adapted to the context in the CoP, the scenario being tested and the tools and services used. The aim of this process is to provide feedback both to developers and to CoP animators to help develop and improve tools and scenarios.

At the same time, WP5 worked on the scenarios to produce a small number of generic scenarios. This work, which was carried out within teams bringing together developers, pedagogues and CoP

---

[66] The WP1 documents talk of "validation" and "validator": These words are problematic in English, conveying a meaning that whatever is being "validated" is done so in comparison with a pre-set standard or goal. This is not the case. The word "authentication", however, refers to the act of establishing or confirming that claims made about whatever is being authenticated are true. This appears more appropriate to us.

animators and mediators, responded to the need to provide scenarios that can be adapted and adopted in many circumstances. The three generic scenarios covered three areas of activity: reification; debate and decide; and knowledge management. During the observation of the use of tools and services mentioned above, work concentrated on those activities to be found in the generic scenarios. In addition, in order to provide workable specifications to developers, the three generic scenarios were further re-worked to provide concrete instances of use.

Internal testing of services (WP4)

During the Palette project, four versions of the CoPe_it! software were produced. Testing of new versions but also of local fixes and incremental developments took place on a separate development server in an on-going process involving all eight members of the team. According to the person interviewed, all aspects of changes and developments were tested. One strategy adopted was pretending to be an end user. In some cases this required creating ad hoc software to simulate extreme cases, in testing performance for example. There would seem to be no written procedure for testing and choices of what to test and how to test it seem to be based on practice and personal experience. Feedback was more structured. Minor problems were dealt with informally. Major problems were discussed in the weekly or twice weekly meetings, and a written record of all actions (including testing) was kept in a shared spreadsheet using colour codes (green – ok; yellow – in progress ; etc…). The spreadsheet was also used to detail forthcoming steps in the work. In the case of complex problems, the results of testing were analysed at team meetings and decisions about re-working the code were taken at the same time.

Market analysis and the state of the art competition (WP4)

Experts are expected to keep abreast of the latest developments in their field. How do they go about collecting such information and how is this information assessed and integrated into current processes? In WP4, research and market watch was an on-going part of the work. Experts read articles by others and attend specialist events. An additional strategy used was to set students who are working on the subject to retrieve information about current developments as part of their research work. In the case of Palette, this work was reinforced by a specific task (T4.7) to catalogue and analyse new trends in specific areas and to reflect on how they could affect future developments of the relevant Palette tools. As in many of the cases we have studied, there was no set procedure for the analysis of the data. 'Experience', or should one say 'flare', enabled experts to identify developments that were potentially interesting. Integration was facilitated by adopting an 'incremental formalisation process' in which simple features (such as joining a community) were developed first and more complex functions added later (such as annotating documents or awarding scores). Decisions about what should or should not be integrated into existing software were discussed in team meetings. Such decisions were conditioned by a number of external factors: the contractual requirements of the Implementation Plan, although there is considerable leeway about how objectives can be reached; the available resources for carrying out additional work; the scientific interest of the developments proposed.

The Crete interviews: challenges in Palette (WP6)

One of the major motivations for the series of five-minute interviews, carried out in Crete during the Palette workshop there, was the need to provide faster evaluation feedback to the steering committee. The interview contained three questions about perceived challenges in the project and what could be done about them including the role the interviewee would play in addressing those challenges. In WP6, it was decided to present the results in the form of quotations structured around a series of key points. There was to be no analysis (other than the regrouping of the answers under certain headings) and no recommendations. It was expected that the SC would draw its own conclusions from the material. The three-page report was sent to Steering Committee members less than a month after the interviews.

The Lyon interviews on formative evaluation (WP6)

During the Lyon Plenary at the end of 2007, all the steering committee members were interviewed about formative evaluation. These interviews sought to gather evidence for the report on the reactions of the Palette project to formative evaluation (D.EVA.04). This deliverable was provided in response to the request during the first review of the Palette project to provide a report on how issues identified in D.EVA.03 had been addressed and how the consortium had reflected on its practices.

The questions asked in Lyon had been prepared by WP6 and sent to SC members prior to the Plenary along with a short summary of feedback from prior evaluations about perceptions of formative

evaluation and its role in the project. The interviews, which were recorded, lasted between twenty and thirty minutes. The interview questions were taken as a starting point in a "client-centred" approach where interviewees' preoccupations were allowed to emerge and develop. Such an approach uncovered a predominant preoccupation with the relationship between evaluation and decision-making, which was not one of the subjects of the initial questions. The results took the form of a six-page document that covered a number of areas: the perceptions of and demands on the evaluation; the relationship between the Participatory Design Methodology (PDM) and the evaluation; the evaluation results as levers for change; and possible pathways forward. The report was submitted to the Steering Committee early 2008.

Future business plan and OS strategy (WP7)

One of the tasks of WP7 (dissemination) is to work on a business plan to ensure the sustainability of the PALETTE tools and services beyond the end of the project. To establish a draft exploitation plan, a questionnaire was drafted and presented to the General Assembly in Lyon at the end of 2007. It was subsequently circulated early January for fine-tuning and the final version was presented to the January 2008 Steering Committee. The questionnaire was then made available online during the month of February to the relevant project partners (service mediators, CoP mediators, and commercial partners). The first part of the questionnaire specifically for commercial partners concerned intentions about using Palette services and critical factors concerning decisions to adopt services. The second part, dedicated to service mediators, gathered information about currently available services in particular forms of availability and dependencies as well as information about future plans including possible strategies, business models and risk analysis. The third and final part for CoP mediators asked about services currently used and projected future use both of Palette and non-Palette services. The report on the outcomes of the questionnaire[67] provides selected details of responses from the partners but proposes no project-level plan or synthesis. A summary of the results were presented during the Review by the European Commission. A final business plan is to be provided before the end of the project. A separate conference was planned to discuss the business plan, but financial considerations will probably require the conference to be integrated in an existing project plenary meeting.

Note that the business plan is linked to an open source strategy elaborated in D.MAN.07[68] and D.MAN.08[69]. The latter states: "*the 'Final Open Source Strategy' deliverable aims at providing indicators in order to ensure the feasibility of the Exploitation Plan/ Business Plan for PALETTE services*"[70]. This document describes work being carried out with the QualiPSo[71] project to audit services developed within PALETTE, to know if their licensing constraints fit the Exploitation Plan defined by the Consortium.

A number of difficulties emerged during this process. Firstly, not all those people addressed by the questionnaire were in a position to make relevant decisions and necessarily had to refer to the appropriate person. Secondly, the interest of the research community for a business plan was limited as it corresponded neither to the culture nor the institutional structures of academics. At the same time, the European Commission put much emphasis on the need to achieve sustainability of the results. This dilemma illustrates the tension between high-level research and market-oriented development, according to the person interviewed.

The RUFDATA Framework

It is made up of seven questions:

1. What are the **Reasons and Purposes** for evaluation?
   These could be planning, managing, learning, developing, accountability
2. What are the **Uses** of your evaluation?

---

[67] D.DIS.07 PALETTE draft exploitation plan, April 2008.

[68] D.MAN.07 Open Source Strategy (draft), November 2007.

[69] D.MAN.08 PALETTE Open Source Strategy, March 2008.

[70] Op. Cit. Pg. 4

[71] http://www.qualipso.org/

They might be providing and learning from embodiments of good practice, staff development, strategic planning, PR, provision of data for management control

3. What are the **Foci** for your evaluations?

   These include the range of activities, aspects, emphasis to be evaluated, they should connect to the priority areas for evaluation.

4. What are the **Data and Evidence** for your evaluations?

   Numerical, qualitative, observational, case accounts

5. Who are **Audience**s for your evaluations?

   Community of practice, commissioners, yourselves

6. What is the **Timing** for your evaluations?

   When should evaluation take place, coincidence with decision-making cycles, life cycle of projects

7. Who is the **Agency** conducting the evaluations?

   Yourselves, external evaluators, combination

The questions used for guidance in the interviews

1. What were the aims of the evaluation?

   e.g. To improve a product or process

2. What activities or objects or objectives did this evaluation focus on?

   e.g. The activities of CoPs in terms of information sharing, …

3. What instruments and/or methods did you use to collect your data?

   e.g. Questionnaire, interviews, observation,…

4. How did you create these instruments or methods?

   Where does the creation of these instruments or methods fit in the sequence of activities in the project and how does it relate to those activities?

5. Where did the parts of the evaluation fit in the sequence of other activities in the project and how did they relate to them?

   This should also include the possible articulation with activities in other WPs.

6. Who carried out this evaluation?

   e.g. The CoP mediators…

7. Who were the results of the evaluation destined for?

   e.g. CoP mediators or developers,…

8. What form(s) did the results take and how did those forms relate to the nature and context of the audience(s) they were to address and to the aims of the evaluation?

9. To what extent and in which ways were other actors associated with the various stages and moments of the evaluation?

   For example: in deciding on and designing the method and tools; in getting people to participate in the evaluation; during the gathering of data; in understanding and analysing the data; in giving form to the results; in validating the results with the community "evaluated"; ...