

# KoWi-Thema

**SEPTEMBER 2005**

## **THE EVALUATION AND SELECTION PROCESS OF COLLABORATIVE PROJECTS**

---

### **SCENARIOS IN VIEW OF FP7 – STRIVING FOR EFFICIENCY**

**DR. DIANE SOMBROWSKI**

---

Herausgegeben von der Koordinierungsstelle EG der  
Wissenschaftsorganisationen, getragen vom Verein zur  
Förderung europäischer und internationaler wissen-  
schaftlicher Zusammenarbeit e.V., dessen Mitglieder  
folgende Organisationen sind: AiF, AvH, DAAD, DFG,  
FhG, HGF, HRK, MPG, WGL und der Stifterverband.

Vorsitzender des Vereins: Dr. Reinhard Grunwald  
(Generalsekretär der DFG); stellv. Vorsitzende:  
Dr. Christiane Ebel-Gabriel (Generalsekretärin der HRK)  
Redaktion: Dr. Diane Sombrowski

**KoWi** Koordinierungsstelle EG der  
Wissenschaftsorganisationen

Godesberger Alle 127  
D - 53175 Bonn

Rue du Trône 98  
B – 1050 Brüssel

Tel.: +49 228 959970  
Fax: +49 228 9599799  
postmaster@kowi.de

Tel.: +32 2 548 0210  
Fax: +32 2 502 7533  
postmaster@kowi.de

---

## EXECUTIVE SUMMARY

---

The study examines the potential development of project evaluation in FP 7. Currently (FP6) great emphasis is put on the ex ante proposal evaluation.

As for the question, which projects should be funded, the peer review based ex ante evaluation presently is the single and widely accepted way in order to prepare funding decisions based on scientific expertise.

However, EU funded projects require more than just excellent proposals. Additional objectives are pursued in a system where policy relevance and outreach are key factors. These key factors have to be paid special attention to.

The study questions whether this has been accomplished in the traditional system of ex-ante evaluation. From that starting point, it will develop three scenarios which may improve the way those key factors are validated in the project evaluation. Two scenarios deal with slight, mainly procedural modifications of the current system. The peer system as is operates certainly stimulates competition for funds, but does it stimulate high achievement? Thus, the baseline here is to look beyond the 'promise of performance' and to stimulate both project flexibility as well as progress. The third scenario therefore suggests splitting evaluation in three categories: An ex-ante part where the basic decision on funding a project is made, a second part which consists of a monitoring of the ongoing project progress and a third part which consists of an ex-post evaluation of both the results and wider effects. That, of course, means shifting distinct evaluation criteria towards a later phase of the evaluation. The shifting has to go along with a new definition of the criteria as they will be used for the evaluation at different times and from different angles.

The study will show that the third scenario is the most appropriate when it comes to the question of an overall evaluation of the success and impact of a project. From that perspective, the rationalised ex ante evaluation is only one - but a very relevant and necessary - step, when it comes to continuous project evaluation. However, only with a combined scheme of ex ante, permanent monitoring until project end and an ex post evaluation the final conclusion can be made, whether the project actually contributes to the goals of European research funding.

## OVERVIEW

SCENARIO I. The scenario exclusively scans the possibilities of reducing the time frame of ex ante evaluation and questions whether there is some leeway to manoeuvre. It is argued that, for several reasons, the time-saving option is not a convincing one with regard to the evaluation phase. Here, the classical trade-off between workflow speediness, cost-effectiveness and high quality procedures clearly appears. Several measures that increase procedural speediness would induce higher personnel costs for the Commission, or would hamper the quality of the present process; especially, the reduction of the time span foreseen for expert selection would have adverse effects on the matching of evaluators and proposals. In this context, less concentration on instrument types and, in addition, 'lighter' contracting and financial processing practices seem to be the right remedy. It is questioned whether a limitation of multiple-stage procedures would impair the evaluation and its flexibility. It is argued that especially the hearing option should be rethought, as

it lacks a clear added value due to its fashion which tends to disproportionately prevent any flexible, open discussion.

SCENARIO II. This scenario looks into options on how to refine the current ex ante evaluation concept. It thus explores several fields of procedural improvements, mainly through refined ways of advanced standardisation. In fact, the proposal evaluation is well-designed and methodologically quite consistent. Nevertheless, the potential of ex ante evaluation can be improved by both stricter, more balanced tasks profiles and guiding principles for consensus and panel discussion - lastly, to enhance the peer review result generation as well as the quality of funding decisions. Special attention is paid to expert recruitment, remote evaluation and the proposal preparation stage. The boundaries, which a systematic usage of two-stage submission encounters, are discussed; especially here, cost neutrality is affected - it is by far no panacea and should be handled with care, especially when it comes to the definition of first stage criteria. Moreover, the merging and cutting down of the operational criteria blocks is discussed. In particular, the handling of the Relevance and Impact criteria should be rethought - it is presumed that a clearer understanding of the criteria would both improve the currently faced interpretation gaps and, in addition, help to stem the tendency towards exhaustive briefing sessions.

SCENARIO III. In contrast, the third scenario highlights a paradigm shift. It is meant to design a scheme in accordance with the life cycle of the individual project; it shall thus generate more flexibility to be in line with the project's progress and governance. Moreover, it shall allow for a less heavy ex ante evaluation in order to award funds in a flexible way based on a more trustful relation between funding body and beneficiaries. It takes the present concept away from the very complex and 'static' ex ante practice that remains an appreciation of various criteria, but refrains from coupling this appraisal with the subsequent project performance. A new approach linking ex ante, interim and ex post assessment implies the definition of new criteria categories and, consequently, the shifting of certain ex ante criteria towards a later assessment phase. The main concern is to come to a scheme that poses the adequate questions at the most appropriate time. So the scenario basically argues for the introduction of distinct evaluation/monitoring dimensions - depending on the project stage. It primarily deals with criteria re-definition, in the sense of a separation between ex ante appraisal criteria, project cycle criteria and project end/ex post criteria, the latter assessing both the direct results and the impacts on the Communities' political and strategic goals (see figure, p.23).

Project impact assessment is an uncertain business, especially at the proposal stage. So it should be decided whether it is sufficient to get a rough ex ante estimation of the likely impacts or whether it would be meaningful to generate a systematic assessment of the outcomes. It is therefore suggested to establish, following the initial appraisal, a continuous project performance monitoring during the project cycle and, finally, an output and impact evaluation at project end and beyond. The suggested movement is towards an engagement of project evaluation with the broader evaluation agenda. The short-term result measurement virtually conflicts with the long-term, yet politically driven European evaluation perspective. In this context, one should strive to observe the project's progress and its needs in order to learn more about the relationships that link the objectives and outputs to wider effects. This is to say that specialist skills will be needed to satisfy the information needs of all stakeholders. The approach is finally meant to foster the visibility of project's outputs and to enable the verification of the project's output additionality, primarily in terms of European added value. In conclusion, further steps are identified (see the listing, p. 25/26).

---

## INTRODUCTION

---

### NEW PATHS IN EX ANTE PROJECT EVALUATION

The purpose of the study is to explore the development potentialities on the basis of the current evaluation culture of the European RTD Framework Programme in order to provide a set of scenarios highlighting possible modifications of the ex ante proposal evaluation in view of the 7<sup>th</sup> Framework Programme and to progressively conclude on how to induce an alternative approach to selecting and evaluating potential collaborative research projects. The study thus addresses the funding mechanisms for the nine thematic areas under the specific programme “Cooperation”. It aims to sketch a practicable evaluation approach that differs from the original Commission’s concept whereby it takes account of the present discussion on procedural simplification and, in general, tries to include actual challenges in the field of RTD evaluation.

Starting point is that the majority of proposers and beneficiaries regard the RTD Framework Programmes an important resource. But at the same time, across the research community, there is still a clear wish to see more stringent science-driven assessment of quality with fewer, less restrictive rules in certain aspects of programme administration.<sup>1</sup> In terms of proposal submission and evaluation, most applicants would much appreciate a new fashion of the concept in order to reduce the bidding costs. Many would like to see the peer review based concept to be fostered primarily to improve quality of the funding decisions and to make them less arbitrary. Beside, most of participants perceive the financial rules and cost models as almost inscrutable and the contract preparation as complicated to an unreasonable degree. Since proposers need to deliver nearly the same information within both the evaluation and contracting stage, just in different formats, they would welcome, if the doubling of data and information requirements at the submission and contracting stages would be eradicated. The same goes for the later cash flow problems which are considered as problems of the programme management culture which is regarded as not being in line with project execution.

Since it is not probable that the numerous complaints are just part of a permanent process of criticism, the streamlining of administrative procedures currently is very high on the agenda<sup>2</sup>. In contrast, the ex ante proposal evaluation is rather subject to an ongoing modification process at the margins. One can observe a slight modification process in the course of the previous Framework Programmes, including a quite distinct incision prior to the 6<sup>th</sup> FP where the operational evaluation criteria were separated from the evaluation manual<sup>3</sup> which comes as a strictly procedural document since then. The revision of the concept was in no case meant as turning point, rather as an internal quality assurance exercise.

But is a ‘pointwise’ modification of administrative workflows and rules reasonable without considering the relation of costs and benefit of the complex ex ante evaluation? DG Research proposes, for instance, an electronic registration desk to avoid any doubling of data requirements. Besides, it is suggested to assure the extended use of two-stage submission procedure.<sup>4</sup> However, since the Commission opts for a strict simplification of the reporting scheme and, in addition, for the establishment of a sound follow up phase, it would be indeed very reasonable to connect the current troubleshooting in terms of user-friendly rationalisation issues with a global rethinking of the project evaluation concept. Thus, the study is not about how the current concept operates, it

is more about how its weaknesses could be overcome and how one could make better use of its strengths. In this context, the study points out potential new paths in project evaluation.

There are two general assumptions the scenarios take into account: (1) the proposal of the 7<sup>th</sup> FP posits ample continuity in terms of priority themes as well as funding instruments, but allows a more flexible use of the latter; (2) the methodology, this is to say project evaluation on the basis of a peer review procedure steered and organised by the Commission is not a moot point. The point at issue is to fathom the mutability of the present system and to discuss what kind of advantages a further development could offer.

---

## HOW ABOUT THE MUTABILITY OF THE SYSTEM?

---

Is the present evaluation concept likely to be profoundly rethought, in other words, what is the main barrier to overcome, if the scope of the concept is to be modified? First to mention there are currently no systematic links established between ex ante and ex post evaluation of projects. Policy relevant S&T research and applications increasingly affect wider society and, as a consequence, this shows the need for a well-designed fashion of project impact assessment, in the sense that “a continuous and multi-layered process that is independent from the interests of single stakeholders”<sup>5</sup> ought to be implemented. The present approach lacks a focus on results and outcome: Even though the proposal evaluation strongly addresses the policy impact of projects, the evaluation and selection phase is completely cut off from subsequent phases.

## THE PHILOSOPHY BEHIND

A number of flaws in the peer review process might arise from the fundamental mindset behind. Peer review is used for advice on which research projects should be funded in the first place. But actually, it is the link between evaluators' advice and final selection that needs to be strengthened and revitalized: Some of the rhetoric concealing problematic shortfalls should be changed: in the first instance, why not saying evaluators' *choice*? The formal selection decision is the Commission's and this overall responsibility cannot be delegated – albeit the weight of the peers' opinion is to be reinforced in order to confine the Commission's margin for manoeuvre after the evaluation by the external experts.<sup>6</sup> There is no valid reason for a however argued prerogative of the administration to re-examine the results of peer review, this is to say, to overrule the evaluators' opinion. Instead there is a clear reason for the supervision of the procedure resulting in the selections. That said, it would be eminently advisable to recognise that the present principle should be reworded: the Commission evaluates the proposals with the help of independent experts. The evaluation by independent experts regarding both content and methodology as the determining factor for the acceptance or rejection of applications is a sound principle that would not thwart the Commission's competencies and management task, if the final *choice* would be with the experts. This slightly changed mindset would help allocating tasks and balancing the task profiles of the involved actors, the Commission and the evaluators. Hence, it would be thoroughly appreciated by the addressees of Community funding.

From this, one can suggest that in case of the present peer review which is presumably the most widely used technique in S&T evaluation there are areas that should be subject to improvement, mainly through rethinking of panel composition, task allocation and decision finding. If we were to start from scratch today to re-design the process that is used to determine how RTD project funding is allocated, it could take different forms of what has evolved in the past decade from pre-existing mechanisms. So sections I and II develop scenarios arising from the call for both proposer-orientation and rationalisation of the evaluation and selection process. They are designed to show varied grades of innovation. Section III acts on the assumption that the main innovation barriers emanate from system-immanent motifs and restrictions; it is meant to highlight synergies with the whole shape of project evaluation follow up, in the sense of a paradigm shift and, therefore, to go beyond the coordinates of the present concept.

---

## THE PEER FACTOR

---

The process to determine how science funding is allocated is a quite unique one on EU level (see figures in the annex). Over the years it tended to become more complex and sophisticated. In principle, the validity of peer review is internationally accepted; the methodology never was seriously questioned. One can even observe the trend towards evidence based policy that means that peer reviews inform decision making across an increasingly wide range of areas. But for all that, a broad discussion about the general value of peer review has intensified so that it might eventually be scrutinised more intensively.<sup>7</sup> If one is to deal with a peer review based evaluation procedure some general constraints need to be taken into account with respect to the long-term, ongoing discussion on the recruitment problem respectively the 'quality' of expert panels.

### PEER REVIEW UNDER SCRUTINY: PROCESS CONSTRAINTS

1. Peer review challenges from the increasingly competitive nature of science, which has implications for panel composition. Even in broad research fields and even highly-recognised evaluators at international level may be affected by funding decisions that allocate resources to their potential rivals. This may influence the execution of the individual assessment and scoring.
2. Another challenge is both multidisciplinary organisation and the arrival of problem-centred research. The latter has led to the thematic organisation of research in politically driven, 'problem-solving' areas (such as e-health, e-commerce) and, hence, has brought together research teams ranging from engineering through to humanities and social sciences. Technology and content span a wide range of disciplines causing the emerging of an advanced interdisciplinarity within projects. This corresponds to the visible trend since the 5<sup>th</sup> Framework Programme putting an increased emphasis to RTD as a means to achieve social goals. That altogether, of course, has implications for the recruitment: a long-established criticism asserts that peers are most often recruited from the centres of disciplines and therefore tend to score down proposals that fall *between* disciplines. The criticism suggests the existence of a

general bias. There is evidence indicating that the problem is most acute where a new discipline is opening up (like in Nest) or combining classical disciplines.<sup>8</sup>

3. The peer review system generally is open to abuse and misconduct due to its reliance on trust: the evaluation relies on mutual trust, as researchers entrust their ideas to peers while peers must presume that project participants are providing trustworthy data. This defines a basic problem not easily to address. However, it has little sense to turn down interdisciplinary proposals on the basis of the comments of just a very small number of peers and apparently, it is indispensable to exhaustively reflect upon the composition and reasonable size of panels types; anyhow, the general panel size should not exceed 10 panel members.

The process constraints mentioned above are probable to emerge within all kinds of systems. Regarding the panel composition and expert recruitment as executed by the Commission there are some basic requirements to be put forward that, mainly because of its contents-orientated emphasis, are even more important to FP7 than to its predecessors:

- In case of evaluation of emerging technologies, the task of panel composition must thoroughly reflect the need for specific expertise within the panels.
- A solution for the rising confidentiality problem with regard to industrially driven research is badly needed.
- Evaluators sometimes find themselves in a predicament, when they assess proposals in their own area of interest, as they may be strengthening their own competitors, if they provide opinion that can be used to improve the proposal. They might be in an even worse predicament in case of large-scale projects: the problem of evaluating NoEs containing per definition most of the experts in the field is to be solved. The establishment of independent NoE-specific committees is the most useful clue.
- Especially for long-term projects there is a need for consistency of the evaluation process - this refers to a certain extent of personal consistency in the *whole* process.
- Experts from the New Member States must be integrated at the best possible rate.
- Gender balance among the peers ought to be enhanced. The Commission's gender mainstreaming policy has been in place for many years, but the percentage of contracted female evaluators stills falls far short of male evaluators.
- Above, all, the evaluation skills need improvement. Since it seems not realistic to establish a harmonized instruction of evaluators without causing major delays in the evaluation process, an elaborated rolling training scheme that independently runs in parallel to the briefing exercises is the best alternative means of improving skills.
- It is important to carry more weight to the individual *comments* - in the consensus and panel discussion evaluators are exposing themselves to debate among colleagues, especially, when they provide 'innovative' comments. It therefore must be assured that the whole range of comments is properly appreciated and, consequently, included in all evaluation documentation (as well in the Evaluation Summary Reports).

In sum, it is worthwhile to strive for a generally strengthened role of evaluators that can be seen as full recognition of expertise. Unless young evaluators improve their proposal writing skills, basically, evaluation should not be seen as a mere exercise in proposal writing - it demands

more than basic knowledge. It should be considered as a specific skill alike as a part of the normal scientific activities and thus more sophisticated training is a precondition for the acceptability of peers. The communication of the shifting role is a task, which should be done jointly and concertedly at national and EU level. Especially, there should be a *systematic* exchange of expertise between funding bodies at all levels.

There is evidence that European high-level scientists tend to stay away from the evaluation task at EU level, which might partially be related to the timing of the evaluation sessions. Of course, the relatively short time period between end of call and the beginning of the sessions impedes the recruitment of highly accepted professionals to a certain degree. But there is possibly a rather intrinsic and profound rationale to the recruitment problem that somehow relates to motivation. In several cases at national level, within the evaluation of high value RTD projects the external peers give an opinion on the proposal, while the ultimate decision is made by a *programme* or *expert committee*. The Commission's concept partly differs from this: The final decision is made by the responsible programme management; it is argued that scientific officers have to consider the budget availability and the budgetary distribution between the disciplines or the strategic objectives of the Work Programmes. The Programme Committees composed of national representatives usually give their opinion or, in some special cases depending on the amount of funds awarded, their approval. But there is another aspect that hampers the evaluation and final selection: The EU project proposal evaluation is a *policy linked* assessment and exactly at that point, there is a flaw in the process. Policy based proposal evaluation sometimes involves "internal evaluators",<sup>9</sup> but here, the concept does not know internal (Commission) evaluators. Nevertheless the evaluation of policy relevance is to be done. The policy impact evaluation notably depends on the instructions given to the evaluators by scientific officers - or it is done by some means or other. Accordingly, it is not surprising, when the scientific community often prevails that the evaluation of politically driven criteria is somehow 'jointly' done. Due to the fact that there is a continuously increasing need for briefing documentation and lessons on the politically inspired criteria, some experienced and well-known evaluators might be deterred.

To clarify and upvalue the role of evaluators it should be set out as a rule that final say in terms of S&T merit of the proposal is with the evaluators (respectively with the consensus meeting and the panel). This is vital to the future evaluation system, as will enormously foster transparency: External experts are primarily responsible for the *scientific rating* of a proposal so that the (extended) panels have the ultimate decision to say no to the S&T merit of the proposals. This type of evaluation presupposes strictly transparent assessment realms.

The call for enhancement of the 'quality' of evaluator's skills is a recurring thought at EU level, but does not entirely hit the mark. In either case, the expert pool with its approximately 35.000 entries gives a wide range of external expertise. Rather, it is about making a good choice and combining the proper expertise. Of course, the key task of ensuring the quality of the evaluation takes time. Therefore, it is not an adequate, capable aspect in context of the simplification debate. In contrast, it implies a basic goal conflict, as the best possible mix of discipline and specialised knowledge requires precise overview of the research circles and networks. The fine tuning usually is a complicated, very costly management task. In case of expert databases where researchers submit their professional details by internet, the idea to improve the quality control system is very often put forward, but difficult to implement. If it is decided not only to rely on experience in past evaluation sessions the most convincing way is to strengthen quality by other means.

Taken together, all these factors show the difficulty to obtain a careful panel composition. New emerging scientific fields, interdisciplinarity, the integration of a higher proportion of experts without evaluation experience at EU level and the scope of big projects altogether reveal that the enhancement of evaluators' skills is a major point and that extended training would be the right option, as mere briefings during the evaluation session are evidently not the adequate measure to cope with all these challenges. A training scheme would be critical to the overall acceptability of evaluators.

---

## I. THE TIME SAVING OPTION

---

This section exclusively concentrates on the time-to-contract indicator. So the essentials remain very near at the FP6 status quo. In particular, the revision of the guidelines would not question the methods of generation of results, nor measures to foster transparency like the recording practice. All indications would stand on increased flexibilisation toward an expansion of leeway of the responsible departments in terms of procedural process flow. The Commission would surely concentrate on the mere procedural matters whereby they would primarily pursue the goal of the security of the bureaucratic operational sequence. Therefore nothing changes in the self-conception of the administration; accordingly, questions of fundamental philosophy are not concerned. The overall authority of the Commission in the area of project evaluation and project selection and its consequences would remain 'untouched'. Thus it is foreseeable that time-consuming, pure administrative bureaucracy is sourced out in any case, since the Commission aspires towards concentrating on the political issues of programme management. The criteria catalogues would be under consideration, but again the criteria modification would be targeted on the regrouping and rewording of the well-known instrument-specific criteria. A shortened catalogue and/or an alternative arrangement are thinkable.

In this setting, solely those process components are regarded, which seem suitable to affect the time-to-contract (ttc, time between closing call and contract signature), which is traditionally used as a monitoring indicator measuring the efficiency of programme management. As FP7 is thought to operate even at a greater financial scale and complaints by both proposers and participants were even more stringent regarding exhaustive administration. To avoid critical delays the streamlining of administrative workflows is very high on the agenda and might also effect the evaluation phase.

However, precisely the time-to-contract-indicator<sup>10</sup> should be put into context: this metric directly interrelates to contract types as well as to plans of budgetary execution that are quite independent of the evaluation scheme - and negotiation, too. The Commission repeatedly argues that consultation of programme committees are large blocks of time and, in fact, comitology affects the ttc, but the Commission usually sets time frames. Above all, the negotiation and contract production stages are exceedingly 'heavy' and time-consuming as FP5 recordings indicate.<sup>11</sup> It is apparent that the whole ttc-time span tends to increase during the execution of FP6. Despite the tremendous work on scaling down the contract types to a single model contract, well in advance to FP7, there repeatedly is the quest for simplifying the negotiations. All stakeholders unanimously assent that administrative procedures and financial rules increasingly seem to disable the management of instruments on either side, the Commission's and the project partner's side, primarily due to the complicated cost models and the ex ante controls.

The remedy for streamlining will be very similar to those during FP6: improved workflow planning, running procedures in parallel, improved documentation and improved communication with participants and a warning system for blockages in negotiations. However, some of these solutions would imply a higher proportion of personnel; for example, it is a question of how many A-staff running a batch of contracts.<sup>12</sup> Under the prevailing circumstances, streamlining could actually be costly. Since it is targeted not to increase the Commission's human resources, there is no sound answer on how to achieve the goal - in the face of the intended scope and scale of FP7. Given the current complexity of the project design there is no comprehensive and striking solution on how to ease the burden of contract preparation and negotiation. The only convincing solution would be less concentration on project type characteristics. Even the

outsourcing of so-called 'upstream tasks', which comprise the receipt and handling of proposals and the financial viability checks is no cure-all. A sound corrective would be a 'lighter' approach in terms of both contracting (introduction of grant agreements) and the financial processing.

How about the effects of the ex ante evaluation process on the whole time span? Given the current concept, the most probable measure to bring about a 'positive' time effect is to change one of the fundamentals of peer review technique: if evaluators would be briefed to early "abandon" proposals which would involve major rearrangements, primarily budgetary cuts or scientific changes, later on this would help avoiding complex negotiations. In the Commission's opinion, this is not desirable, since external experts should normally neither be instructed on how to form their judgement on budget allocation nor to sort out specific kind of proposals in order to simplify negotiations. This would be a mode of rigor that would unduly impair integrity of the outcome of the external assessment, since it influences the opinion of the peers.

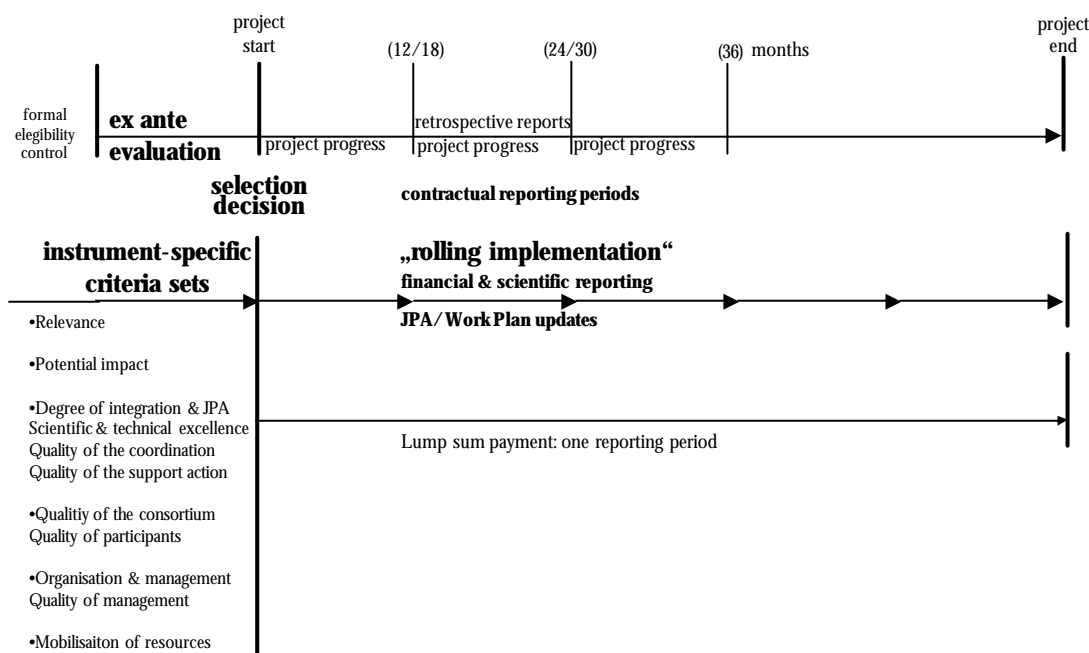
Concerning the procedural strand of the proposal evaluation it is probable to cut down the core of the process, notably the consensus and panel stages. The evaluators running the individual assessment are anyway confronted with an immense time pressure. Evaluators are asked to stay in Brussels on average one week. In order to reduce this burden the panel stage could be skipped, as it actually is a doubling of the consensus discussion, of course, at a higher scale. What is the benefit of the panel stage? One could argue that there should be other means of synthesising the consensus results than a sometimes doubled panel stage (distinction between the 'panel' and the subsequent 'extended' or 'synthesis panel'). The Commission would hold against that the panel meeting is a crucial part, as it allows reaching higher resolution for the purpose of improving decisions through the re-examination of the consensus conclusions. A second option would be the withdrawal of the hearing option actually set up to reassure the opinions formed at previous evaluation stages. Indeed, the added value of this component is to be questioned: the hearing rules confine the meeting to an unreasonable extent, since they prevent any dynamic discussion. Hence, it should be discussed, whether the hearing could be further developed towards a less restricted meeting allowing interactive debate. Alternatively, pre-defined questions should be answered by the consortia in writing.

In addition, there would be another potential component that could be investigated with respect to time saving. Even though the briefing sessions are highly appreciated by most evaluators one could argue that the instructions should be cut down to a minimum, since they tend to be overloaded. On the other hand, the importance of specific briefing raises the bigger and more complex project types become. Recent trends in the proposal evaluation concept even tend to raise the necessity of highly differentiated briefing sessions – not only due to complex project types, but also due to complicated instrument-adapted criteria. According to FP6 Independent Observers, at several stages a detailed briefing is essential: i.e. before the remote stage, before and during consensus discussion as well as before the extended panel meeting.<sup>13</sup> It follows that neither the procedural fashion of the evaluation nor the briefing by Commission staff would be subject to the time saving exercise. On the other hand, it is clear that the tendency towards excessive briefings could be prevented by a lighter instrument design and a reduced criteria system. As mentioned, the key task of combining the skills of external experts helps ensuring the quality of the evaluation. Therefore, the task of panel composition is not an adequate aspect in context of the simplification debate. In contrast, it implies a basic goal conflict between speediness, cost-effectiveness and good quality evaluation. For instance, Independent Observers suggest considering a minimum of a 2-month-period between the call deadline and the beginning of the evaluation. Panel composition is rather a field for further improvement, since there is a lack of rules dealing with expert recruitment and the combination of skills and special expertise.

The overall picture here is to show that the time saving option is not very commendable. The mere concentration on cutting short the original evaluation sessions is not a reasonable way to reduce time-to-contract delays unless the fundamentals are rethought. Other procedural phases are more likely to be changed. So within a progressive approach the main aim is not to turn 2.5 months into 1.5, it is not all about the mere economy of time, the aim is rather to make turning the time span to a solid process. In sum, the scenario shows that there is just a restricted leeway to improve the timing. The options to really reduce the time span without hampering quality are sparse and, in fact, they would negatively affect the procedural flexibility, one of the main advantages of the peer review method – except for the hearing that suffers from its inflexibility; it should be redesigned towards an open debate exercise or substituted by a lighter writing exercise. Basically, the major field for ttc-improvement is the duration of contract preparation.

## II. THE REFINEMENT OF STATUS QUO: THE RATIONALISATION OPTION

### The present model: separated ex ante evaluation



The second scenario accounts for the Commission's position on the above mentioned classical trade off: According to the Commission, stronger emphasis is put on the quality of the whole selection process and, in particular, the safeguarding of the evaluation rather than on the duration – thus the Commission is highly reluctant to examining the appreciated proposal assessment in the light of the pure economy of time. Many persons in charge as well as practitioners tend to regard the development of the common evaluation concept as having reached a plateau, more than ever, since FP6 Independent Observers indeed point out some

shortcomings, but altogether attest a high quality standard. In this scenario the refinement of procedural aspects is thus at the centre and, to a lesser extent, the improvement of the evaluation results correlating with the improvement of the quality of decisions through criteria rewording. The principle of an instrument-adapted criteria design would be under scrutiny by no means. Again the Commission would strive for a modified presentation mode of criteria: possibly there would be efforts made to shorten the catalogues towards having less, preferably equally stable criteria and, very probably, again there would be some re-arrangement of criteria blocks in order to integrate (split up and subsume) the Relevance criterion. The outsourcing would comprise a wide range of clearly administrative tasks, while the political tasks ('downstream tasks' like expert recruitment, panel composition, evaluator briefing, contract preparation) would remain with the Commission, even though this trend, namely the differentiation between downstream and upstream tasks, will complicate the execution of the evaluations – the outsourcing should be directed towards a logical aggregation of interrelated workflows. However, since Commission staff traditionally has a very 'intimate' relationship to political aspects of RTD funding, the most likely reasoning behind is the intention to stay very closely to the strategically important projects.

Currently, the Commission is trying to find ways on how to enhance the efficiency in the management, thus how to reduce the unsustainable legal and financial requirements and the time-to-contract. To that effect, the modification of the financial regulation and especially the cost models are centred. However, it is a moot point whether the subjects under discussion in view of the administrative environment are effectual and sufficient.

The streamlining and simplification efforts primarily concentrate on contractual and financial issues. So there are some general *assumptions* that are taken for granted; the scenario stands for the measures planned by the involved directorates in terms of simplification, especially the following: reduction of a priori controls to an indispensable minimum; the simplification of required project data at the proposal stage as well as at the contract preparation stage; the setting up of a centralised registration system and database; the establishment of a new on-line tool replacing the Evaluation Service Provider; the re-definition of eligible costs ("used-approach") together with the definition of a less intricate cost model towards a thoroughly revised grant to the budget; a fairly extended use of lump sum financing for justified costs; and a particular model for NoEs including lump sum payment warranting a completely autonomous spending of the flat rate within the scope of the joint programme of activities.

From a practical point of view, there are three aspects that would predominantly be subject to the exercise of procedural refinement: the proposal preparation stage respectively the reduction of submission requirements, especially concerning project management; remote evaluation (for individual and consensus stages); and the two-stage submission procedure. A higher degree of refinement could only emanate from an advanced standardisation therapy. As regards evaluation criteria the handling of the Relevance criterion is to be re-discussed.

PROPOSAL SUBMISSION. The proposal preparation would not be high on the agenda. The Commission often stated that the amount of details required at the submission stage would be reasonable compared with other funding bodies (NSF); the presently required data are seen as indispensable for statistics and contract preparation – so it is not probable that the Commission will demand for less detail in the administrative forms (FP6 form A). But in the case of Integrated Instruments respectively its successor, the heaviness of the process might be released through slight changes of submission form B requirements respectively the decrease of requirements addressing management data: the material being related with the micro-management of the project could be cut down to a minimum. If it is the intention to stringently reduce the

paperwork being produced with regard to project management, it should be consequently communicated that visualisation like Pert, Gantt-Charts and others are generally not expected and do not influence the assessment and scoring in any way. This measure, of course, can only succeed, if it would be communicated in a reliable, thus harmonious way across the whole programme. As a consequence, this would ease the burden of putting together an application and would help to concentrate on critical requirements otherwise overlooked. The same goes for evaluators: the additional value of rather secondary managerial information is questionable in relation to the restricted time frame. In general, it is important to revise the application forms with the aim to avoid requirements referring to collateral information. As to the set up of a pre-registration, it should be explicitly stated that a compulsory mechanism will be introduced, since this positively affects the composition of panels as well as data management in general.

REMOTE EVALUATION AND EXPERT RECRUITMENT. Contrary to the preparation issue, remote evaluation is dealt with thoroughly. It is seen as a measure to possibly facilitate the recruitment of relevant experts. It is seemingly very difficult to recruit high-level experts to spend one week in Brussels. The individual evaluation is a suitable task for remote evaluation and even some remote consensus would be generally thinkable. For that purpose, the Commission aims to spend money for a new tool especially designed to handle remote evaluation and replacing the Evaluation Service Provider. If the individual reading is meant to assess the quality of proposals, remote evaluation could give an answer to the question whether a project is worthwhile funding or not. In contrast, the consensus and panel stages are a comparative, interactive assessment exercise and therefore take place in a central place – so it is not advisable to regularly organise remote discussions. Moreover, one should not forget that remote assessment can not be used, where commercial confidentiality is necessary – and this goes for most EU projects.

The effect of remote evaluation on expert recruitment is not predictable; the effect might be a positive one, but not a compelling one. Hence obvious disadvantages should be weighed: first of all, an extended use of remote contravenes the nature of the result generation based on 'corrective' discussion and consensus. In this context, there is an additional drawback: remote experts who are not present in Brussels will not have any possibility to defend their conclusions; in these cases, equal treatment of judgement could be undermined.

MULTIPLE-STAGE PROCEDURES. During FP6 a wide range of varieties in terms of the two-stage submission has been evolving. Across the Thematic Priority managements, a consistently limited proposals size (20 pages) was introduced. Nevertheless, there was no consistency as regards the application of criteria at the first stage: in fact, the use of a *limited* set of criteria was rather uncommon at the outline stage, whereas short proposals often were evaluated against the same criteria as the full proposals. As a consequence, proposers were asked to address all criteria, just less detailed – this, of course, does not result in a competition of ideas. When a reduced set of criteria was used, yet the outline assessment was not a competition of research ideas, since the proposers had to provide evidence to Relevance and Potential impact; thus, they had largely to expand upon the European dimension, European added value and the potential advance in socio-economic terms. The latter very clearly unveils that there is an intrinsic weighting of criteria within the system, which was not in line with the original sense of the criteria catalogues. Following the recommendations of the *Marimon*-Panel and the *Five-Year Assessment*-Panel 2004 the two-stage submission procedure will be further developed to partially relieve over-subscription and improve proposal preparation conditions. But the extended use of outline and full proposal stages is by far not a panacea, since it affects the evaluation period and is thus critical to timing. Additionally, it doubles the incoming applications and increases administrative costs. Alike remote evaluation, the multiple-stage are time-consuming and cost-intensive for both the

Commission and the applicants. Unless a common understanding of first stage criteria is agreed, there is no certainty that an outline proposal will cost less time and money.

On the other hand, two stages are suitable where the research topic is defined quite open and a very broad interest across European researchers is therefore expected. For instance, when the call text is less prescriptive regarding the S&T area supported; this mainly goes for areas that are drawn from the edges of cross-cutting disciplines. Furthermore, the multiple-stage procedures are reasonable for complex projects types that need a very long time for preparation and consortia building purposes. In this case, the costs for intensive proposal preparation might erode the value of funding so that the reduction of bidding costs through an outline stage is very reasonable. Another advantage is that the first submission stage could be initialised before the consortia building is finalised. In that light, it is most important for the reliability of the decision-making at the outline stage that the basis of few criteria is well defined and clearly understood. A more progressive two-stage procedure should be able to solely rely on the criterion of scientific excellence at the first stage, in the sense that a mere competition of ideas would take place, while in a second (and third) step secondary criteria would intertwine: this is basically feasible in case of new emerging areas. In sum, a selective use of multiple stage procedures should be envisaged, such that the procedure properly reflects both the inputs associated with proposal preparation and the research activity involved - it would rather be a retrograde step to unnecessarily complicate the submission process through the general use of the two-stage procedure.

ADVANCED STANDARDISATION. In terms of fine tuning and refining the evaluation process the most likely efforts would mainly concentrate on enhancing the conditions of work, which is to be highly appreciated. But are the current plans solely addressing the costumer's sufficiently strong? When aspiring towards an advanced degree of rationalisation, there should be some more efforts made particularly with regard to strengthening the links between peer review results and the Commission's final selection. Whenever a quality process is meant to be reviewed, the strict documentation of results and the rules for opinion generation at all evaluation stages should implicitly be reconsidered too, and finally, it should be clearly set out which parts of the documentation are coercively integrated into the Evaluation Summary Report.

Advanced standardisation would clearly mean defined procedural standards, especially dealing with the critical stages of the process, namely discussion and scoring at the consensus and at the panel stages. According to FP7 Independent Observers the preparation of extended panels is a crucial task, as there is lot of pressure towards the generation of the final evaluation outcome. Therefore specific rules on how to handle the comparative exercise at the highest level and on how to resolve disagreement and reach conclusion are highly necessary. The discussion activities do not only need to be organised but also to be regulated by norms that aim at constraining actions of individuals and attaching value to the decision finally taken. At the same time this would help defining the important *task profiles*: the definition of a profile for Commission moderators chairing the meetings would be important, since there is a need for a more consistent handling of consensus discussions. If the Evaluation Guidelines were less vague, especially concerning the role of Commission moderators, the need for clear additional rules could be reduced. In this context, one can question the necessity of call specific guidance notes for evaluators that, in fact, differ just in detail, mainly due to inconsistent use of criteria weightings and procedural differences (appointment of panel chair, scoring). Moreover, more consistent, stricter rules could help attenuating the present trend towards an excessively differentiated, multiple-step briefing scheme.

There is lack of publication requirements aiming at clear transparency standards: particularly, the regular publication of the recommendations of Independent Observer should be introduced. In general, the peer review process should be regularly overviewed by involving external advice from evaluation professionals. To this end, one could be inspired by international good practice; for instance, the establishment of an advisory body would significantly upgrade transparency.<sup>14</sup>

CRITERIA DESIGN. The Relevance criterion is recognised as most sensitive to divergent interpretations. According to FP6 Independent Observers, the criterion is the one posing the most problems to evaluators.<sup>15</sup> Bearing this in mind, the Commission consequently reconsiders how to handle Relevance. There are three most likely options to lower the heaviness of the process: Firstly, it could be decided that the intrinsic quality of the proposal is expected to match with the Work Programme; this would lead to a new criterion 'type', since the control would be primarily executed by Commission staff and later on *affirmed* by the external peers. Secondly, the criterion could be reworded. Thirdly, Relevance could be split up and incorporated in other criteria. The latter, a mere regrouping of the current criteria blocks, seems highly preferred. However, the solution of merging, for instance, the *Impact* and *Relevance* blocks together is not commendable, since this would help concealing a critical criterion, but not clarifying the contents, nor improving the understanding. It is quite foreseeable that the Commission will opt for this solution: The texts asking for justification on Relevance and Potential impact (Guide for Proposers) reveals that both criteria include the EU dimension respectively the European added value in a very similar way. These considerable overlaps should be avoided by separation and clarification. Otherwise one of them could be completely sorted out. Therefore, the handling of politics-driven criteria is an issue worth for exploring. This remains a major task in criteria design.

For the majority of clients it has been harder to participate in the New Instruments. Even for FP4/FP5 experienced organisations it was complicated to become involved. As a result many preferred to take a rather passive role in developing proposals or in executing projects. The funding schemes might run the risk of losing contributions from experienced smaller groups. In this light, it is a crucial task to establish a clearer common understanding of what constitutes quality and excellence at the EU level. This would be especially helpful to the East European research community. Similarly, the absence of a strong and unambiguous definition of the European Added Value (EAV) is a difficulty hampering the evaluation exercise and causing irregularities and unsteadiness in the interpretation and implementation of the criteria. Closer integration of research policies requires mutual understanding of what constitutes European Added Value in contrast to National Added Value. All parties involved need to comprehend this at a very early stage of the programme execution.

## SYNOPSIS AND CONCLUSION

In this setting, it is only the IT infrastructure that would be widely revised in order to improve remote evaluation. Neither the criteria design nor the panel composition would be thoroughly revised in a first step, even though each area should be subject to in-depth consideration. The free choice of the funding mechanisms would be the most important novelty. The proposers will not directly be affected by the fine tuning conducted in this setting - apart from slight changes that might be introduced in proposal formats at the preparation stage. Nevertheless, applicants would benefit from a general refinement of proposal evaluation. But the scope of these modifications can be questioned. However, at a *higher* degree of refinement, there are some challenges suggested that would help fostering transparency and providing a higher innovation level. The revision of documentation and publication requirements as well as the

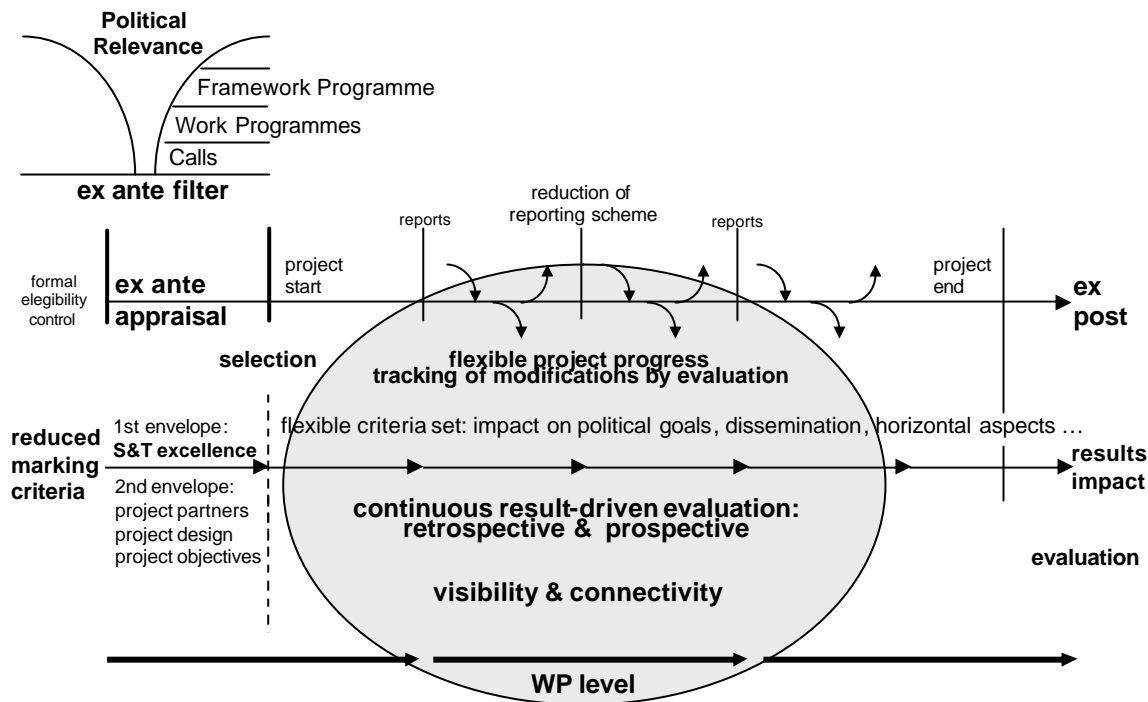
generation of evaluation results would yield advanced standardisation. In this context, it is suggested to regularly publicise the recommendations of Independent Observers. The Commission might feel motivated to seek advice from evaluation experts during the quality assurance process by establishing an oversight group involved in the procedures that pertain to peer review. The characteristics of panels should be spelled out as well as the tasks, rights and restrictions of evaluators. At the same time limitations of panel composition should be clearly communicated. Most importantly, the design and common understanding of operational criteria should be high on the agenda. Only the reduction and complete reformulation would assist evaluators in interpreting and applying the criteria. Accordingly, this would improve both the steadiness of criteria and the quality of results. As positive by-effects, this would allow for the adaptation of the submission requirements (form b) and for a rationalisation of the briefings sessions that were permanently growing over the years, mostly due to vast criteria explanation.

Regarding two-stage submission one can observe much variety across the FP6 Thematic Priorities. The common ground is the standardised proposal size at the outline stage. But in terms of the identification of criteria at the first stage the programme administrations seem to disagree. The Commission should be reluctant to use the complete range of criteria in the first stage, since this tends to water down the competition of ideas. Above all, the Commission should be very clear about the Excellence/Quality criterion as first choice.

The setting shows that the Evaluation Service Provider would be renewed in order to introduce a regular remote evaluation tool and improve expert recruitment. It should though be considered that a more intensive use of remote evaluation would contravene the basically corrective discussion mode of consensus and panel evaluation. Instead, problems of equal treatment of judgements arise. Remote evaluation is not the clue to expert recruitment. Alternatively, the pre-registration option should be discussed. The introduction of a regular mechanism would provide sound and timely information on the mix of disciplines required for the assessment. This would give more time for the selection of experts and would improve panel composition. But apparently, this measure would fall far short, since the choice of experts is a very time-consuming and difficult task. A more complex and open-minded approach, including networking with national agencies, is to be discussed – the identification of experts should go beyond the limitations of the expert database. Permanent structures (for a detailed view see below, section 3) respectively the involvement of evaluation professionals would eminently assist in identifying ‘right’ experts.

### III. BEYOND STATUS QUO: STRIVING FOR EFFICIENCY

#### New approach: designing a sound follow-up



It may be that the question is not one of pure simplification and/or refinement of the status quo to overcome procedural obstacles, rather one of a shift of focus in terms of proposal evaluation procedure towards a broader perspective - for the ultimate purposes of improving decisions as well as the feedback to stakeholders. Secondly, the challenges raised by new implementation instruments may take away the evaluation from the basic form of peer review, the practice may change, most notably the evaluation culture. Altogether this requires fundamental changes that go beyond the advanced refinement therapy as presented in section II.

In this scenario, the starting point is to examine on how to react more closely to the *process* of the project approach. The efforts therefore revolve around recognising profoundly the needs of past high-potential clients in order to retain them and, of course, to gain new high-potential clients. This means to put more emphasis on providing support through evaluation to both the projects as well as to the programme management units during projects life time.

The scenario accounts for the externalisation of administrative tasks as well as of the financial processing and combines it with the need for a continuous evaluation scheme during project life cycle that can adapt more flexibly to the needs of the single project whereby it gives necessary leeway to the Commission and allows concentrating on core tasks connected with its overall competence. Further on, the scenario relates to the simplification measures as outlined in the previous section (see assumptions, p. 13), most notably the reduction of a priori checks and the redefinition of the reporting scheme.

CONCENTRATION ON EX POST. The scenario primarily refers to a statement that was communicated by the Commission with regard to implementation modes and management tools for FP7. There is clear message on how the Commission intends to achieve the simplification message and it can be recognised as a re-orientation message:

*“The general approach will be one of trust towards the researchers. Specific measures will include providing more autonomy to the consortia, simplifying the Commission’s controls and concentrating them on verification of results ex post rather than on ex ante controls, reducing requests for information and reporting.”<sup>16</sup>*

A result-driven ex post control - what does that imply? In the first instance, it brings up the urging issue of financial controls. But beyond that, from a broader perspective, the change mentioned foreshadows an opening up of the borderline between ex ante and follow up. To meet this intrinsic challenge, it is necessary to re-think the whole proposal evaluation, not only the financial strand. First of all, one needs to note that ex post evaluation has its roots in accountability. Similarly, RTD evaluation is linked with accountability, but at the same time there is an ample interest in *strategic* issues and in assessing effectiveness. Hence it is meant to assess both the possible long-term impacts and the final output in relation to the input. In terms of the evaluation exercise, the message quoted above includes a link to accountability respectively to new public management and the related emphasis on checking performance. Accountability generally leads to the use of audit-type methods, sometimes extended to include research inputs and outputs. The evaluation type generating feedback to strategic changes examines the appropriateness of policy goals and of actual and possible directions of RTD and its system. RTD evaluation for decision support involves judgement by peers/evaluators which is transformed into scores and produces information which is available in the system. The ‘evidence’ derived from the information generally could be taken up at various levels.<sup>17</sup>

In the original concept, the Commission uses the external ex ante information/opinion once: the selection decisions are based upon the evaluator’s advice and recommendations are then used as an input to contract composition. Logically consistent, the information could also contribute to resolve the question of result-driven project monitoring by orienting ex ante evaluation towards a later continuous checking, better: tracing of performance and linking proposal assessment with the following ongoing project progress, this is to say the initialised project evaluation could be extended in order to enable a sound follow up in accordance with the Commission’s intention to foster the follow up exercise. This provides a sound *steering* option, as explained above. The running of a continuous respectively regular project evaluation by external evaluators would have relevant advantages:

- first of all, it allows to rationalise the separated, complex ex ante proposal evaluation since it allows for a shifting of ex ante evaluation issues which are then more closely linked to project execution;
- it gives consistency and strength to the result-orientated follow up process;
- it allows a clear differentiation between core criteria on which the ex ante exercise should indispensably concentrate (compare graphs, p. 11, 17) and a second envelope of criteria which are both less important with regard to the funding decision and which are difficult to assess *prospectively* and, therefore, can cause incoherent interpretation and use in the ex ante evaluation; it would be very meaningful to assess the projects against the second envelope in the course of project implementation;

- the permanent involvement of professionals enhances credibility and
- last but not least it automatically results in both a stable project end and ex post assessment, as a certain procedural pressure emanates from the regular tracking scheme. This kind of automatism coordinated in cooperation with the Commission actually serves both the connectivity to broader evaluation agenda and the Commission's high interest in the overall strategic supervision.

So far, ex post evaluation was sometimes used as a format at project level, as it provides better information about track record and can serve as input to other evaluation levels with the ultimate aim of organisational and policy learning. But seemingly, current ex post evaluation includes monitoring rather than evaluation activities. It is important to bear in mind a basic determination: Audit gives evidence to regularity and legality of expenditures, monitoring refers to information on management performance, while evaluation relates to information (learning) and judgement (opinion) covering inputs and outputs during project implementation phase as well as outcomes, this is to say long-term effects. Whenever the scenario refers to the continuous approach, here 'monitoring' is not only used in the restricted sense of information on management performance, but with respect to a broader project *assessment* against distinct criteria.

In sum, the approach is meant to 'ease' and rationalise the very complex ex ante evaluation concept struggling with oversubscription and process flaws, especially the criteria system which obviously asks too much interpretation work from its users. The aim is to assure process quality by reducing the work overload of evaluators on the one hand. On the other hand, the reduced complexity of the criteria system is an appropriate formula permitting a straightforward use of criteria sets and averting vast interpretation activities. Evaluators are apparently facing a tremendous time pressure, since proposals tend to become more and more voluminous and descriptive, and there is no way to avoid this trend given the steadily increasing competition. This trend should be faced by setting up a new set of criteria clearly showing which core components become crucial to the implementation. In this context, it should be considered to quit the instruments-adapted criteria sets, as FP7 implementation is meant to be less instruments-driven. At the same time a reduced set would help evaluators to concentrate on core requirements and characteristics; accordingly, it would provide more flexibility to the mere assessment exercise and would enhance the scoring. As by-effect, the new criteria design would help curtail the steady growing of the briefing documentation.

On the other hand, the approach is meant to help establishing a robust, high-quality follow up-process in which the cultures of retrospective appraisal and forward-looking advice on the implementation could be reasonably combined. A professional approach, in particular through the involvement of evaluation experts, is getting increasingly important, since the follow up becomes weightier in the context of an ex post achievement-related system. The project size in FP6 is on average three times as large as in its predecessors and as a consequence thereof the single project has become considerably crucial to the whole programme. In terms of the review strand the regular evaluation/tracking would eminently provide evidence for the rolling payment scheme. Especially in case of a new NoE-specific funding model as well as for lump sum payment, this could be a reasonable and simple backing for the steering exercise. The regular evaluation does of course not completely replace the audits, but it is meant to give *flexibility* to the reporting scheme following the Commission's intention to cut down the overcharged annex requirements of the periodical reports. In this regard, a strictly harmonized reporting scheme would not very useful. In contrast, as a prerequisite, a project specific, reduced reporting proving leeway to the project management is badly needed.

At the same time the *retrospective* element of evaluation on the basis of project-driven criteria would enable a tracking of project results towards the pre-established objectives of the single project and, would serve the call for concentration on high-potential projects. The continuous evaluation driven by experts would provide robust information on which a decision on the cessation of project could be based. In addition, the *prospective* element of evaluation on the basis of thematically driven criteria would form valuable advice on flexible work plan modification; this exercise should not be linked to Commission's formal approval any longer. Firstly, it is arguably apparent that, in a way, the approach would create a new type of evaluator: the evaluation during project cycle implies a peer as a 'companion' instead of a 'controller'. To a certain extent, the evaluator would act as a mediator – between project management and Commission. Moreover, it may be a key to a better balance between the role of the programme administration and the role of external experts. The nature of the evaluator's opinion would intrinsically be upgraded. But altogether it does not necessarily mean that personnel continuity is absolutely mandatory. To prevent the continuous scheme from 'miscast' the establishment of a permanent *panel structure* is suggested, in which an automatic personal exchange is system immanent (see below). Secondly, the combined approach suggests an elaborated criteria system including a careful categorisation aiming at a kind of 'criteria ranking' in the *ex ante* phase and a flexible use of criteria to a certain extent in the interim and *ex post* phases. Giving greater weight and more *visibility* to project's results and wider outcome means to track project progress more *closely*, but less uniformly. The continuous evaluation offers a solution to the call for more tracking and evaluation of new research findings so to increase the likelihood that the work might be exploited to a greater extent.

Aside, the scenario might help providing a reasonable model that, due to its fashion, suggests a re-balancing and separation of the two main blocks of proposal evaluation and selection, namely the assessment of a set of scientific and project design related issues and the financial assessment. To that effect, the simplification of the financial strand is an implication. To put it in other words, the continuous evaluation scheme could be regarded widely separated from prior and subsequent financial checks which could consequently be accomplished by the executive agency. If a bunch of several duties and responsibilities dealing with programme implementation would be assigned to an executive agency, indeed it would be meaningful not to merge both ERC-related management and other parts of the Framework Programme in order to foster the coherence of workflow. Of course, a complete decoupling of the financial assessment is not feasible, since the equipment fundamentally differs between disciplines.

The model faces two main challenges: the establishment of core panels and the renewal of the criteria system. The following paragraphs shall highlight the need for establishing a new 'status' to evaluators and for a sophisticated differentiation of criteria.

(1) THE ESTABLISHMENT OF CORE PANELS. The execution of projects includes the need to internally manage the monitoring of progress and result. This is the interface where the external on-going evaluation could get into the act, but in low-key mode, since a more flexible and autonomous project management shall be ensured. So as to facilitate the acceptance of external experts one of the cornerstones here is the establishment of a *permanent* panel structure, not actually as a long standing committee, but as thematic core groups though allowing for personnel rotation - to avoid inward-looking dynasties. This feature would assure a considerable *consistency* as a means of quality assurance, since it tends to improve the mode of generating evaluation outcomes – it keeps consistency in decisions, most notably for long-term projects. Through this, the steering by the Commission would be evidently upgraded, since there is a profound knowledge of project progress. The assessment that could be in line with milestones would offer

a professional justification of project performance and output – it would thus provide a solid reasoning the Commission could act upon in case of finalisation of projects ahead of schedule.

Peer review sessions are likely to become ever more difficult to set up due to the high number of interdisciplinary combinations (see above, Peer review under scrutiny, p. 5). It seems that the bureaucratic recruitment exercise is beyond its capabilities to accommodate. Of course, interdisciplinarity makes panel composition even more difficult. This problem is likely to be most obvious in case of evaluations of NoEs. As a consequence, specific NoE panels ought to be composed. It would be very straightforward to disaggregate problem-centred component parts of a complex and highly multidisciplinary proposal for a separate ex ante appraisal executed by extra panels, such that one could think of a model that matches special expertise at different levels; this would argue for a multi-step procedure at the full proposal evaluation stage. Interdisciplinarity clearly indicates the limitations of a permanent core panel, and, in this context, it becomes obvious that rules for the mix of expertise would be extremely valuable. So highly recognized experts in certain fields could do the assessment of specific questions remotely, while a standing panel composed of external experts and Commission staff could prepare the decision based on both their own assessment and the specific points raised by additional peers.

Recently Independent Observers have opted for the establishment of a more sophisticated briefing scheme that should include several steps of specific briefings prior to the evaluation stages, especially prior to the case of a plenary (extended) panel. This obviously would place a heavy burden on Commission staff as well on evaluators. Due to time restrictions a multiple step briefing is not feasible. To solve this problem, a detailed training scheme should be introduced, ideally including a specific training for inexperienced evaluators. The remedy is to have thematically oriented core groups so that new evaluators learn from experienced ones. Like that, a slight renewal of the expert base would be initialised.

(2) RE-DESIGNING EVALUATION CRITERIA. Past experience shows that keeping an eye on a robust selection filter is one of the most exigent tasks. In advance to FP6, the criteria were adapted to reflect the changing needs of the New Instruments, resulting in a set of instruments-adapted catalogues. The extent to which the instruments-specific criteria system has been a move towards a more reliable and comprehensible one is very controversial. It is an understandable approach, as the issues faced in the evaluation of the New Instruments have considerably moved away from those that had been covered by the previous peer review for more than a decade. Or put it in another way, with the ERA concept the Commission has initialised a policy shift from the pathway traditionally occupied. At operational level, the supporting of the collaborative research was accordingly accompanied by a diversification of the funding mechanisms. So the criteria catalogues differ, as they are to correspond to a certain project type. But they do not differ profoundly: they all comprise the same buzz words – whether terminology changes sometimes – and these buzz words obviously shall form the instrument's core. The present pitfall is the merging of the scientific idea with other, quite vague political/strategic aspects. So it is not astonishing that the criteria lists are under scrutiny again previous to FP7. But what gives reasonable evidence that the same modification mode like before yields criteria which will perform strongly? The rewording, splitting up and regrouping of criteria blocks are probably not the best way to achieve strong and standardized criteria - there is no sound reasoning behind.

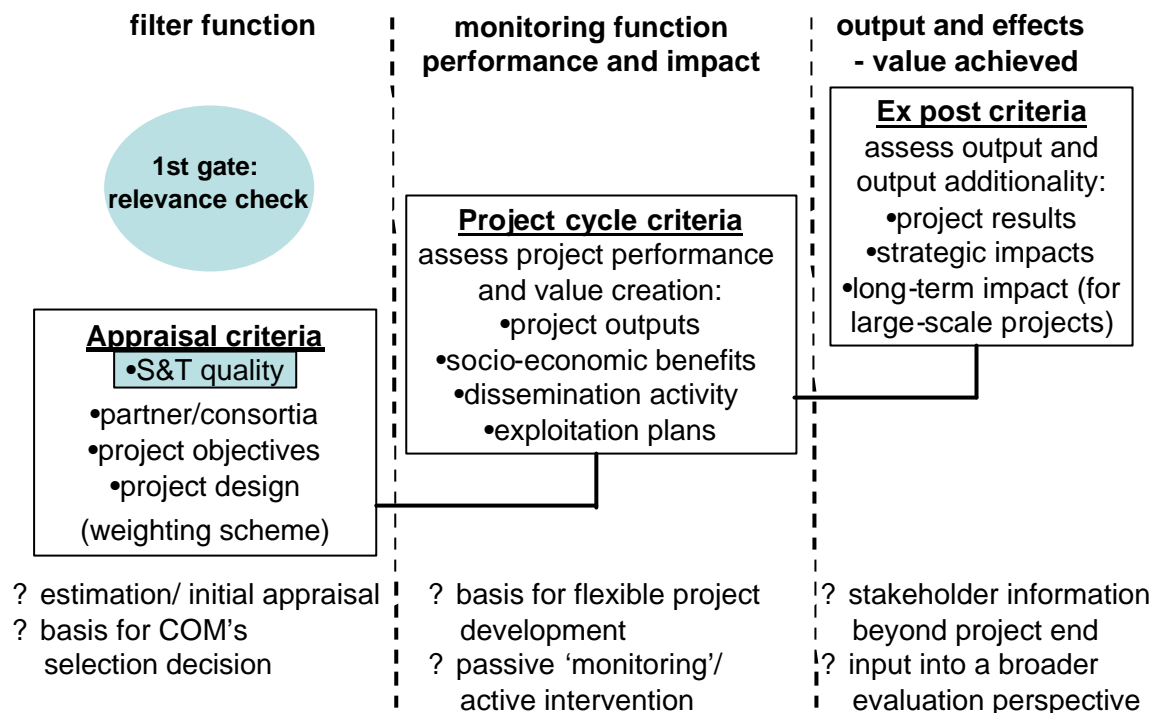
The handling of criteria, especially the Relevance Criterion, regularly poses problems to the Commission group in charge of evaluation practice and it will do so in future. After the FP6 preparation phase, in which Relevance was discussed at great length, it still remains an open question (see scenario II, p. 15). Thus, the handling of Relevance and various Impact dimensions

is a key to the variety of criteria – in general, there is a significant diversity of criteria that was animadverted by the last Five-Year Assessment Panel.<sup>18</sup>

Why not investigating the *categories* of criteria? – In order to explore the criteria that guide the selection of the projects eligible for funding. A selection filter can be both a criterion as well as a pre-evaluation filter (eligibility criterion). Thus, is it possible to differentiate between types of criteria, the one more suitable for the first round of *formal* selection (like relevance), the other clearly the *core* selection criteria, namely scientific excellence/quality and finally, *secondary* (not: auxiliary!) criteria. If it is realisable to find a plausible designation and differentiation, it would be reasonable to establish a simple graduation that helps clearly distinguishing the most important filter with a high degree of precision and makes selection more steady and reliable. This includes that criteria currently used as *ex ante* dimension, such as ‘horizontal aspects’ and similar impact dimensions, would be cut off from the critical *ex ante* phase and shifted to the ongoing/interim and *ex post* phases; this goes for all sort of socio-economic aspects,<sup>19</sup> as ‘strategic impacts’ are arguably dependent on the performance mode; but very typically they emerge in the long run, especially ‘wider impacts’. Hence performance indicators should not be too simplistic, since results can move away from the objectives originally set, and long-term impact should be carefully traced, since unpredictable benefits (effects) and *output additionality* can arise later on. One should bear in mind that much of what S&T provide is intrinsic and remains intangibles. These wider effects are quasi not measurable in advance - “the objectives and the expectations from the evaluation effort have far outreached the pool of measures.”<sup>20</sup> Therefore, more emphasis should be given to the choice of filter criteria that serve as *estimation* criteria. In order to reduce interpretation gaps, the less stringent *ex ante* criteria should be sorted out and shifted.

Due to the fact that RTD - no matter if open-ended or mission-related is addressed - is not a routine activity, it rather is a difficult business to derive *specifiable* results. Thus simple performance indicators should be handled with care, and the tracing of RTD impacts in a broader context is a tremendously complicating task. There obviously is a conceptual problem relating to the simple fact that results are often unpredictable and long-term. This consideration suggests a very careful discussion on the appropriateness of evaluation criteria and a separation between *ex ante*, interim and *ex post* evaluation criteria (see the proposed evaluation dimensions below).

### 3 evaluation dimensions – criteria shifting



In the discussion on suitable ex ante, interim and ex post assessment criteria the following 'dimensions' should be primarily pinpointed:

- The increased emphasis on realisation of socio-economic impacts from research – the trend towards including socio-economic issues within the scope of programme evaluations may be the 'extension' of the relevant group of stakeholders, but is not meaningful at the ex ante project stage. Hence, other means of integrating the relevant groups should be examined at project level. The strategic socio-economic impact assessment should be shifted towards the ongoing tracing.
- The merging of political and horizontal issues with criteria of a completely different nature – the trend towards increasingly highlighting politically driven criteria should be stopped. The diversity and incoherence of political criteria is very high. Above all, they should be shifted, because they could be handled more cogent in the follow up. If the Commission's wishes to retain the political issues within the ex ante filter criteria, the contents should be accentuated of what should be achieved in a harmonized way. In general, impact measures should be clarified.
- The Relevance - a kind of political criterion, too - seems to be the least understood, since it is not clear at all on what it does refer - case-by case: does it address the topic covered, the novelty of the project approach, the scientific relevance, the societal relevance, the impact on training of project staff ...? If a proposal covers more than one topics of the Work Programme, how to score it as it might be more relevant than a proposal that covers just one topic? Would it be feasible to translate the criterion into a "go" or "no-go" criterion or would it be possible to transform it into a mere

selection criterion being checked prior to the evaluation – as a prerequisite, this would imply a more stringent wording in the programmes.

- The management and resources criteria should be cut down to a minimum because it is in doubt, if evaluators are well prepared to suggest budget and cost reductions in every case. In particular, it is to be questioned whether the suggested cost reductions, which are part of the subsequent contract negotiations, are justified in a robust way. The proposed partner/consortia criterion in connection with the project design criterion gives sufficient evidence on the fashion of the management – and correspondingly, a solid project design, which conforms with the pre-defined objectives, implies good project governance. Above all, it should be discussed whether a lighter budget appreciation, in the sense of a plausibility check in relation to the project's overall design, would be advisable.

## SYNOPSIS AND CONCLUSION

In sum, the follow up phase would benefit from a systematic connection with the ex ante evaluation phase. Insights into the causal relations between project inputs and outputs are the essential prerequisite of project evaluation. Hence, it is concluded that proper and consistent ways of combining ex ante and continuous project evaluation ought to be investigated. Besides, this would entail an automatic and improved ex post project evaluation.

The main conclusion is that, regarding its position within the overall evaluation process including several evaluation methods, there should be a convergence between all project phases due to the fact that the combination of ex ante, interim and ex post assessment allows a development of the present criteria towards a reduced system of clear and stringent criteria facilitating the selection and decision-making and enhancing the subsequent assessment of performance and impact. The study therefore suggests the shifting of certain criteria dimensions towards the ongoing exercise instead of a mere regrouping of present criteria aspects.

Further on, it is concluded that the assessment of both performance *and* results is needed during projects' life span to enhance visibility as well as interaction with other evaluation levels. One would thus expect a combination of both monitoring ensuring if the project is on track and evaluation that is mainly about the impact of a project. In this respect, the setting of project aims and objectives is highly important from the beginning; this exercise should be scrutinised by the ex ante evaluators, as the knowledge obtained from the continuous/ex post evaluation is basically valid with reference to those original objectives. By this means the whole process would be bound to the initial assessment of the project input and the originally defined aims and objectives of the project, while the implementation plans can be revised during its life span. To this end, a strong correlation between ex ante and ongoing project evaluation should be established.

A crucial question of the follow up phase is who takes which responsibilities during the implementation of projects and, in this context, who receives the results of the continuous evaluation. Of course, the same goes for the ex post phase in which the impact of the project is predominantly assessed. But during life span this question is more vital: it should be clearly set out that the Commission bases its interim decisions concerning project modifications respectively project finalisation on the results of the coherent evaluation scheme.

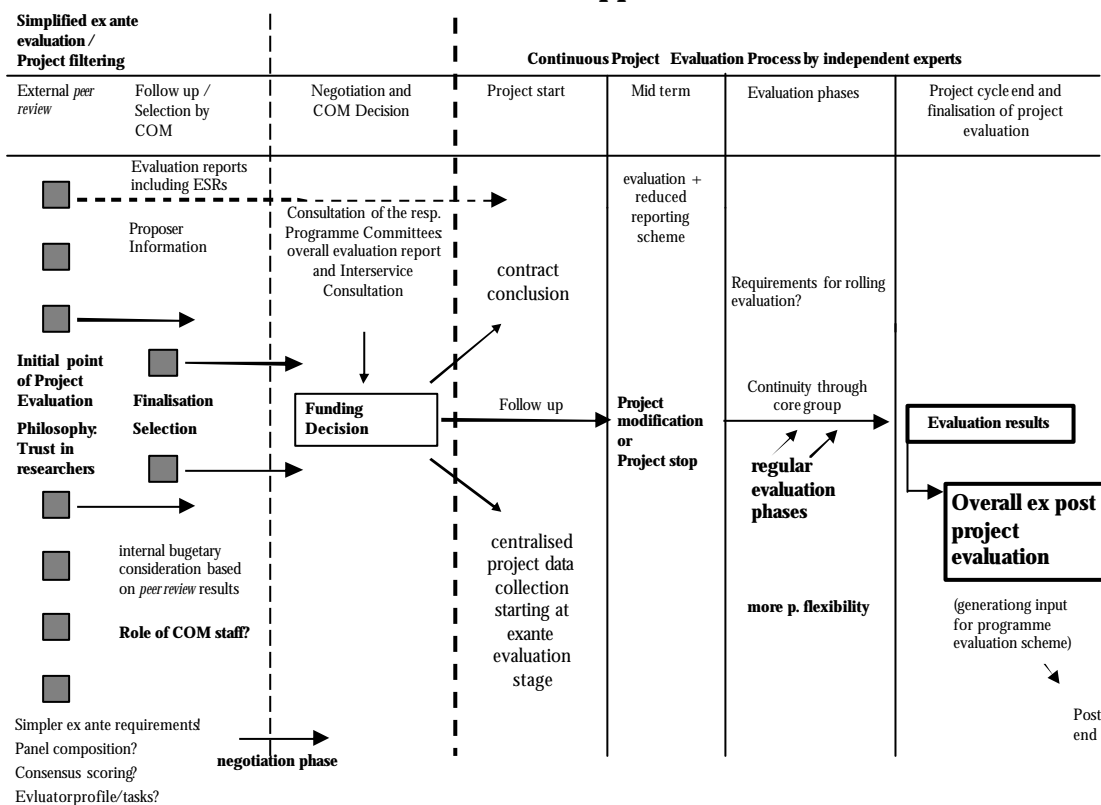
## **FURTHER STEPS TOWARDS A NEW APPROACH**

Winding up, the following steps are suggested:

- (1) Within the present concept ex ante evaluation is restricted to the go or no-go decision making. In the relation to the huge amount of project information being gathered at that stage, the scope of the exercise is to be reconsidered: it should be additionally seen as a starting point of continuous project follow up and thus being linked with interim and ex post evaluation. There should be a strong link between initial ex ante project appraisal, selection and subsequent continuous project evaluation. This link is to be established through a panel structure. Additionally, terminology should be rigorously clarified: project follow up includes both partly monitoring, but mainly evaluation. External evaluation is meant to largely substitute the current reporting scheme.
- (2) The establishment of permanent thematically driven core panels is to be combined with a training scheme conducted by professional evaluators. Particularly for NoE, it makes sense to establish a specific panel procedure, since this funding mechanism widely differs from other project types. It should be concluded that both the specific concept of 'excellence' as well as the measuring of wider project impacts (outcome) require a sophisticated permanent scheme for validating.
- (3) Thus all project evaluation phases should be properly linked together in order to interact by means of a reduction of ex ante appraisal criteria and a shifting of specific criteria dimensions towards the monitoring during project life cycle. At the same time, this requires a reshaping of the present follow up phase: on the assumption that the reporting scheme is meant to be cut down to a minimum, this implies that especially the extensive S&T reporting requirements could be cut down to an absolute minimum, since the continuous evaluation provides a robust information on which the responsible programme management can rely.
- (4) It should be carefully considered if the Commission prefers an active or a rather passive steering model; they could opt either for the option of interventions or of no interventions. If it would be decided that continuous evaluation is not only meant to observe project progress, the type and scope of intervention would have to be defined.
- (5) The *active* continuous evaluation suggests that the evaluators interact and intervene with the steering. This leads to solid interventions, in the sense that project modifications are managed in a more result-driven, thus flexible way – independent of the rigid monitoring. The finalisation during project cycle is adequately justified by experienced evaluators.
- (6) Since the continuous evaluation gives high-quality feedback, for instance, aligned with pre-defined milestones, it should be examined if the rigid financial payment schedule could be decoupled.

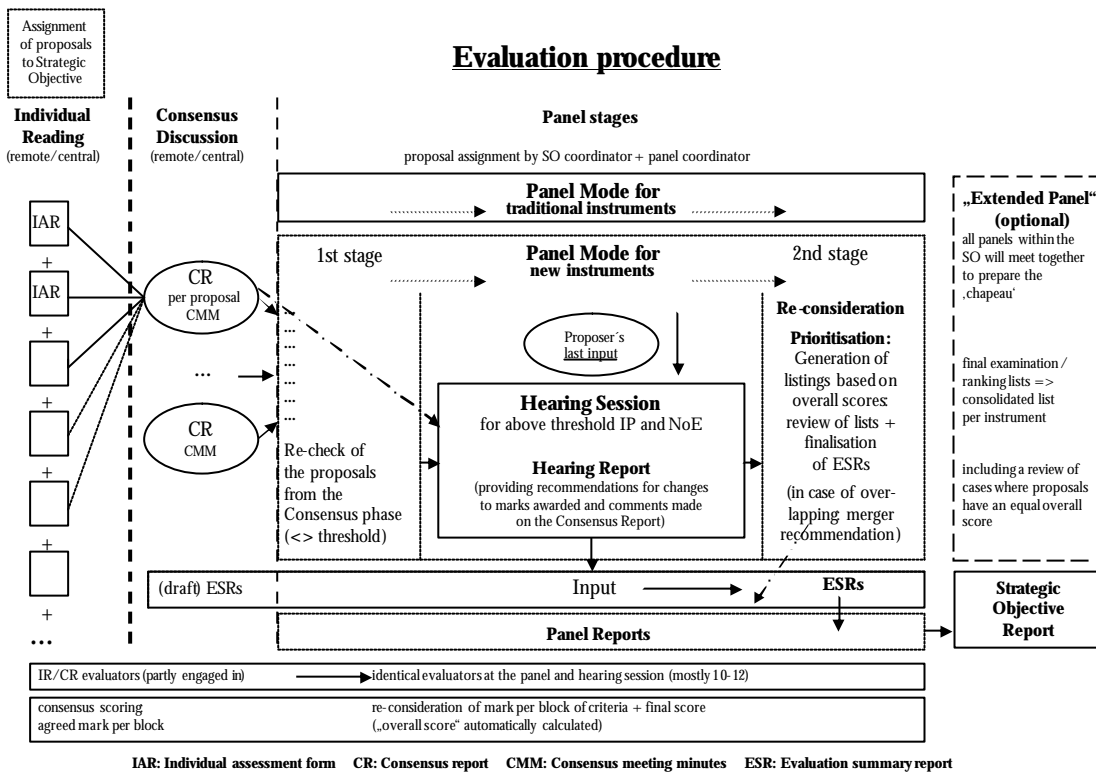
- (7) The reduction of the reporting requirements should correlate with the questions posed by the continuous project evaluation. Hence the evaluation scheme should be built in line with the projects needs, in parallel, the regular reporting scheme is to be discontinued.
- (8) An in-depth revision of the usefulness of the current ex ante criteria and, consequently, a re-definition would necessarily take place to result in a more stable system. In terms of ex ante marking criteria, there is a need of stringent criteria reduction. In parallel, S&T excellence ought to be clearly set out as core criterion; in a second step, the pool of secondary criteria ought to be identified on which the evaluators pose their opinion in a more flexible way. In this context, it is necessary to examine, if there is a common understanding of the definition of 'scientific excellence/ quality'. Only those criteria which are sufficiently robust to be scored should be chosen for the ex ante filter. Criteria that are not suitable for an ex ante appraisal should not be meant to undergo the scoring; in fact, they are to be measured and assessed during project cycle and afterwards.
- (9) This kind of graduation of criteria is indispensable for a consistent follow up tracking progress, results and, finally, the outcome. The present system shall provide balanced criteria block; but this seems not to be viable since the lists are adapted to the instruments. In contrast, emphasis should be placed on core criteria that concentrate on content, not on project type.

### Overview: New approach

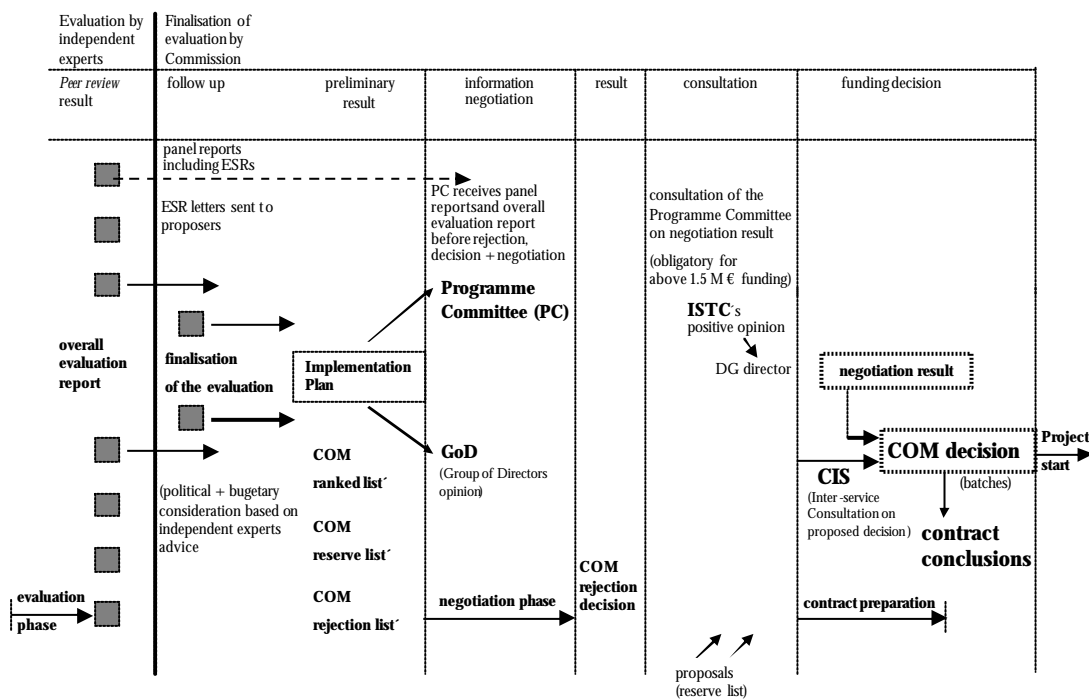


## ANNEX

### OVERVIEW: THE CURRENT PROCEDURE



### Selection and negotiation phases



## **REFERENCES**

Erik Arnold (2001) *Illegitimi nil carborundum...* , PMAC Meeting Brussels, 10 December 2001

European Commission (2004), Evaluation of the effectiveness of the New Instruments of Framework Programme VI. 21 June 2004.

European Commission (2004), Five-Year Assessment of the European Union Research Framework Programmes 1999-2003, 15 December 2004.

European Commission (2005), Proposal for the Council and European Parliament decisions on the 7<sup>th</sup> Framework Programme (EC and Euratom), COM (2005) 119 final.

European Commission (2005), Impact Assessment and Ex Ante Evaluation, SEC (2005) 430/3.

European Commission (2005), Commission Staff Working Document, Simplification in the 7<sup>th</sup> Framework Programme, SEC (2005) 431/3.

European Commission (2004), Guidelines on Proposal Evaluation and Selection Procedures, Amended version, 18.05.2004, COM C/2004/1855.

David H. Guston (2003), The expanding role of peer review processes the United States, in: Philip Shapira, Stefan Kuhlman, Learning from Science and Technology Policy Evaluation, Experiences from the United States and Europe.

Luke Georghiou (2003) Evaluation of research and innovation policy in Europe – new policies, new framework?, in: Philip Shapira, Stefan Kuhlman, Learning from Science and Technology Policy Evaluation, Experiences from the United States and Europe.

Stevan Harnad (1998), The invisible hand of peer review, in: Nature web matters, 5 November 1998 ([www.nature.com/nature/webmatters/invisible/invisble.html](http://www.nature.com/nature/webmatters/invisible/invisble.html))

Macey Henneberg (1997), Peer review: the Holy Office of modern science, in: natural Science, Vol. 1, Article 2, 1997 (<http://naturalscience.com>)

Sean Mc Carthy (2001), The Evaluator/Agency Relationship, Workshop Report from PMAC Workshop held in Brussels 8 November 2001.

Meeting of Independent Observers of the 1<sup>st</sup> round of FP6 evaluation sessions, summary report, rapporteur Günter von Sengbusch.

Meeting of Independent Observers of FP6 Calls, Lessons learnt from observer's findings and discussion on questionnaire survey, rapporteur Mariane Frenay, December 2004.

National Science Foundation, NSF (2004), Grant Proposal Guide, July 2004.

Office of Science and Technology OST (2004), The Impact of the EU Framework Programmes in the UK, An independent report for the Office of Science and Technology by Technopolis, July 2004.

Robin Peek (2003), Could peer review be wrong?, in: Information today Vol. 20 No. 4

Research Councils UK (2005), Evaluation: Practical Guidelines.

Arie Rip (2003), Societal challenges for R&D evaluation, in: Philip Shapira, Stefan Kuhlman, Learning from Science and Technology Policy Evaluation, Experiences from the United States and Europe.

Diane Sombrowski (2005), Inpotevaluation in der europäischen Forschungspolitik, Gestaltung und Implementation der Projektauswahl durch die EU-Kommission.

## **NOTES**

---

<sup>1</sup> For an overview, see for example the report on UK participants for the Office of Science and Technology OST (2004)

<sup>2</sup> Commission staff working document, Simplification in the 7<sup>th</sup> Framework Programme

<sup>3</sup> European Commission (2004) Guidelines on Proposal Evaluation and Selection Procedures, amended version

<sup>4</sup> Commission staff working document, Simplification in the 7<sup>th</sup> Framework Programme, p. 4

<sup>5</sup> RTD Evaluation Toolbox, p. XIV

<sup>6</sup> As regards the status of evaluators' advice, in the Guide for Proposers of the Food Quality and Safety priority one can find the following: "Experts conclusions are examined by the Commission with the support of the Food Quality and Safety Programme Committee ..." (call FP6-2004-FOOD-3-A); this virtually is contradictory to the position communicated by the Commission.

<sup>7</sup> Particularly in the U.K. it was recently discussed, whether there is a need to examine peer review itself; see Peek (2003)

<sup>8</sup> For a detailed discussion see Georghiou (2003), p. 71

<sup>9</sup> Mc Carthy (2001), p. 5

<sup>10</sup> For FP5 the Commission recorded an average of 268 working days time-to-contract across specific programmes and types of activity. Equitably it must be added that there was a considerable spread between programmes: sometimes there the difference was around 3.5 months or even more than 4 months depending on the nature and complexity of the projects as well as on the number of partners involved.

<sup>11</sup> During FP5 the period from end of negotiations to the formal selection decision took an average of 66 days, according to the Commission to a great extent due to the consultation of the Programme Committee. For the *whole* block spanning the negotiation and contract preparation stages a period of 8 months was recorded for the first contracts signed, a period of more than 13 months for the last contracts signed. The partial time span for mere contract signature took 6 months for 90% of all contracts; the remaining 10 % of 'slow projects' took up considerably more time for contract signature – according to the Commission mostly due to slow return of signed contracts and complex negotiations affected by multiple partners, inexperienced contractors as well as significant changes suggested by external evaluators.

<sup>12</sup> In FP5 most staff dealt with 12 – 20 contracts each.

<sup>13</sup> Meeting of Independent Observers of FP6 calls (2004)

<sup>14</sup> One example is the Peer Review Advisory Committee established by the U.S. National Institutes of Health.

<sup>15</sup> Meeting of Independent Observers of FP6 calls (2004)

<sup>16</sup> Communication from the Commission, Building ERA of knowledge for growth, p. 12

<sup>17</sup> For a full discussion on the three functions of RTD evaluations and its intrinsic tensions and pressures see Arie Rip (2003)

<sup>18</sup> 5-Year Assessment of the EU Research Framework Programmes 1999-2003, p. 10

<sup>19</sup> Socio-economic criteria reflect the project's effects beyond the direct benefits, this is to say the project's impacts on the EC's policy goals; thus, socio-economic aspects are strategic/political criteria. The so-called horizontal aspects (synergies with education policy; the engagement with the wider public to raise public awareness and to explore wider societal implication; gender mainstreaming activities) are just a part of the whole range of socio-economic criteria.

<sup>20</sup> Eliezer Geisler (1999): The metrics of technology evaluation: where we stand and where we should go from here, p. 4