

Position Paper on Horizon 2020:

Future evaluation of the Framework Programme



January 2017

Technopolis Group Position Paper on Horizon 2020: Future evaluation of the Framework Programme

Author: Technopolis Group

● ● ● Key messages

1. The European Commission has a solid policy management basis for demonstrating the impact of its research and innovation policies. H2020 and future framework programmes should be at the forefront of practice in monitoring, evaluation and impact assessment.
2. Good design and evaluation of the FP depends not only on administrative procedure but crucially on the Commission's ability to function as a knowledge organisation
3. The Commission should ensure that officers working with design and evaluation are well connected to internal and external knowledge about the FP and to systematic knowledge about research and innovation
4. A knowledge organisation needs a way to maintain an organised memory – both in terms of a knowledge base and in the form of human knowledge and experience. In the case of FP evaluation, it needs a judicious combination of expertise in the domains within which the FP works, evaluation expertise and expertise in the individual tools and techniques used in performing evaluations. There is no one-size fits all methods mix
5. Evaluation should be based on a mixture of qualitative and quantitative methods. The precise mix used in any particular evaluation should depend upon the character of the intervention and of the evaluation questions, and will vary across the FP. Methods innovation is important but should be undertaken in a systematic and quality-controlled manner
6. In recent years the EC has made great strides in collecting and standardising project monitoring data and in generating more access to them. Further efforts are necessary in order to make them more appropriate, relevant and openly available to evaluators and others
7. To grasp the opportunity to be at the forefront of evaluation approaches in research and innovation policy the European Commission should boost its own evaluation competences and improve its governance structure

● ● ● Introduction

Technopolis has supported the European Commission (EC) in the design and evaluation of the Framework Programme (FP) for some twenty years. We have written large numbers of evaluation and impact assessments at the Framework Programme, sub-programme, cross-programme and instrument levels. We have analysed the FP from a long-term perspective and from the perspective of almost all the individual member states that have undertaken studies. We have participated – and are participating – in FP evaluation committees and panels and are the authors of scientific papers on the subject of the FP, evaluation and policy design.

The scope and purposes of the FP have evolved since its launch in the mid 1980s and inevitably, influencing the practice as well as the importance of FP evaluation, increasing the importance of evaluation in providing the information for learning, design and accountability.

The FP has grown in complexity and ambition, now structuring its support around the societal challenges and creating a platform for a more integrated approach to research and innovation (R&I) in Europe. It is an integral part of European research and innovation strategy and also influences R&I policies in the EU Member States. The continuing economic crisis in Europe has implied an increasing emphasis by European policy makers on the use of evaluation results for accountability purposes, underpinning the arguments for funding the FP.

Our remarks in this paper are intended to support the EC in the continuous improvement of its FP evaluation practices.

● ● ● Building on the existing potential for evidence based policy making

The European Commission has a solid policy management basis for demonstrating the impact of its research and innovation policies. H2020 and future framework programmes should be at the forefront of practice in monitoring, evaluation and impact assessment.

We agree strongly with the recommendation of the FP7 Evaluation Expert Group¹ that “considering that the Framework Programmes have consistently been the third largest budget of the European Union, a strategic and professional monitoring and evaluation system is required that increases transparency and serves as a comprehensive and trusted source of evidence-based decision making.” The European Commission has considerable achievements to its credit in establishing evaluation as routine and for improving the data sets available for monitoring during FP7 and Horizon 2020.

The European Commission’s Better Regulation Guidelines² stipulate that ex-ante impact assessment, monitoring, evaluation and impact assessment are integrated parts of the policy cycle, which requires careful forward planning and the integration of evaluation, starting in the preparation phase of new policy measures. The Guidelines set out the principles of systematic evaluation and provide tools for both ex-ante and ex-post evaluation. They stress the importance of evaluation to support both accountability and evidence-based policymaking. We welcome this systematic approach to public management.

Nevertheless, while this provides a basis, this systematic approach is not implemented systematically across the H2020 relevant policy directorates. As the FP7 Evaluation Expert Group also observed, the management of the evaluation actions for the FP is still organised in a very fragmented fashion. The Group observed signs that evaluations do not fulfil their potential as instruments for facilitating continuous learning processes and the development of a solid strategic intelligence.

The Better Regulation Guidelines are developed for policy domains that span the entire breadth of the EC. Their use needs to be tailored to the specific characteristics of R&I policy and informed by experience of the caveats of evaluating public investments in science, research and innovation. While some degree of common structure is needed to compare across programmes and instruments, the evaluation approach should flexibly meet the needs of the individual intervention.

In our view the EC’s achievements in building up monitoring and evaluation procedures and activities, more sophisticated than in many of the Member States. H2020 provides the opportunity to use these in a more strategic way to support future evidence-based decision making.

Adapt evaluation approaches to ● ● ● how H2020 actually works

Good design and evaluation of the FP depends not only on administrative procedure but crucially on the Commission's ability to function as a knowledge organisation

The Commission has chosen to focus its programme design and evaluation effort using 'intervention logic'. This involves systematic analysis of the problems that imply an intervention is needed, the objectives it should have, what activities it should perform and how the sequence of events expected to result from the activities will correct the situation. This provides an explicit 'theory' of intervention that can be used *ex ante* to explore its likely results and impact and *ex post* to evaluate whether these were in fact realised. Strengths of the approach are that it unites design and evaluation via a common logic and guides the search for indicators to help in monitoring and evaluation. A weakness is that it focuses on **intended** impacts, at the risk of ignoring **unintended** effects, whether these are positive or negative. One of several reasons for using independent evaluators is that they are better placed to identify unexpected effects that those responsible for the programme.

The Commission should continue to use intervention logics in FP design and evaluation but also pay particular attention to unintended effects.

The success of all interventions turns out to depend upon external factors, so evaluation can often show that an intervention was **necessary** in order to achieve an impact but it normally shows that the intervention alone was not **sufficient** to do so. The underlying problem is that we lack a proper philosophy of causation when we try to estimate impacts. One result of this is that economic impact estimates are often far-fetched, since they are calculated as if the intervention were responsible for the entire impact (which itself is something hard to measure). An alternative approach – which the Commission has itself tried in relation to the FP – is to identify impacts to which the FP has **contributed**. But a weakness of contribution analysis is that – at least as currently practised – it is massively labour intensive.

Unfortunately, the use of intervention logics does not mitigate what the evaluation community refers to as the 'attribution problem', namely deciding what proportion of the apparent impacts of an intervention should be attributed to it. For example, while an intervention could be evaluated as being responsible for the emergence and growth of a new class of products, that growth will also depend on a host of other factors such as the provision of investment capital, the availability of workforce skills, entrepreneurship, the willingness of users to adopt the new product and so on.

It is not clear that there is a 'right' way to estimate impacts. The Commission could usefully encourage the use of diverse approaches and further innovation in method.

Designing a good intervention logic is only possible if the programme designer or evaluator has or can access a deep understanding of the FP: its policy context, history and not least the lessons that can be drawn from experience. Some of this knowledge exists in the form of past evaluations and studies of the FP. Other parts are in the wider set of knowledge generated by those who work in the technical domains addressed by the FP as well as the multidisciplinary community that does research on research and innovation. Yet more lives in the heads of people in the Commission (and, indeed, at national level), who have worked with the FP for a long time. It is important that the EC masters these sets of knowledge to make the Commission a **learning organisation**, systematically using experience to improve its performance.

The Commission should ensure that officers working with design and evaluation are well connected to internal and external knowledge about the FP and to systematic knowledge about research and innovation.

During most of the FP's history, evaluations have focused on the near-term outputs and results of the programme, giving at best a partial view of its benefits and costs. More recently, greater account has begun to be taken of its longer-term and systemic effects, which turn out to be significant not only in the technical sphere but also in areas like agenda setting, helping establish new fields, disciplines and types of business, building communities and affecting international treaties and the Ozone layer.³ Many of these longer term impacts are hard to discern when evaluation focuses closely on the present policy cycle or the recent past.

Routine evaluation of the FP and its programmes needs to be accompanied by other evaluative work decoupled from programmes and programming cycles to explore longer term as well as unintended effects.

● ● ● Invest in expertise

A knowledge organisation needs a way to maintain an organised memory – both in terms of a knowledge base and in the form of human knowledge and experience. In the case of FP evaluation, it needs a judicious combination of expertise in the domains within which the FP works, evaluation expertise and expertise in the individual tools and techniques used in performing evaluations. There is no one size fits all methods mix.

A learning organisation needs **memory**. Part of this is the maintenance of knowledge bases or data bases – whether of FP-specific studies, evaluations and data or of relevant knowledge that accumulates in peer reviewed journals and in the bigger set of studies and evaluations of research and innovation produced in the course of developing policy at the national and other levels. An equally crucial part is the maintenance of community of people within the Commission who work with this knowledge over longer periods. Without this, the Commission will become prone to ‘organisational forgetting’, the repetition of previous mistakes and a failure to generalise from experience. A third element is the ability to maintain a dialogue about design and evaluation with the considerable community of DGs and agencies involved, supporting its needs for knowledge and expertise.

Since the time of the third Five-Year Assessment (led by Erkki Ormola), DG-R&I has maintained an open knowledge base of evaluations and other reports about the FP, which has been valuable both in the Commissions own evaluations and in the work of others. It is, however, incomplete and there is little systematic access to data about the FP, so there is scope for improvement.

FP design and evaluation expertise is fragmented across several DGs in the Commission, even if DG R&I plays a particularly important role. Personnel continuity in the area has been limited. The Commission should create a common or shared group with a cross-DG remit in FP evaluation and the ability to retain staff over longer periods. It should have professional strength in evaluation but also in the wider set of knowledge needed to support good design and evaluation.

The fact that design and evaluation of the FP requires knowledge that goes a long way beyond administrative process also has important implications for the teams which perform evaluations – whether they are panels convened internally by the Commission or external contractors. These teams need to have, or to be able to access, expertise on evaluation and evaluation design, the FP and other relevant parts of research and innovation policy, the scientific and technological domains involved and the specific evaluation tools employed (such as interviews, surveys, scientometrics, text mining, peer review, etc). The balance among these forms of expertise and the choice whether they are provided within the individual evaluation team or are accessed by the team depends upon the nature and objectives of the individual evaluation.

The Commission should pay attention to achieving the right balance of skills in evaluation teams, especially in the case of panels where it plays a greater role in designing and steering the evaluation and shaping the evaluation team than with external contractors. Crucially, every team should have capabilities in evaluation as a process, and not only in the technical domain being evaluated or in individual analysis techniques that may be of relevance to the evaluation.

● ● ● Methods mix and innovation

Evaluation should be based on a mixture of qualitative and quantitative methods. The precise mix used in any particular evaluation should depend upon the character of the intervention and of the evaluation questions, and will vary across the FP. Methods innovation is important but should be undertaken in a systematic and quality-controlled manner.

Good and robust evaluation requires the use of mixed methods combined with triangulation among the results that they produce. The growing ambitions, increasing thematic diversity and the proliferation of instruments in the FP over the years increase the complexity of evaluation needs and have profound effects on the way evaluations are conducted and there is a need, now more than ever, to adjust evaluation methods and indicators to reflect the nature and expected impacts of the various interventions.

The Commission should continue to support the use a mixed-methods approaches and the triangulation of evidence from different methods to produce robust evaluations.

The FP7 expert group rightly argued for more coordination and standardisation in the collection of some information to support evaluation across the FP. However, a standardised approach will not capture the effects of instruments as varied as the ERA-NETs (and the related Joint Programme Initiatives), Pre-commercial procurement, SME instrument, Marie Curie mobility grants, to name a few.

Austerity pressures since the financial crisis, growing activism and Euroscepticism among the Member States and even the key performance indicators associated with Europe 2020 together trigger a drive towards demanding quantitative accounts of the benefits of the FP. Numeric indicators such as numbers of jobs created or new economic activity are easier than qualitative analysis to communicate to politicians and to a wider audience. However, a bias towards approaches focused on estimating those indicators also causes pressure to produce estimates that are unreliable or not credible. It also creates a risk of underestimating the value provided by qualitative methods, which not only provide valuable conceptual and contextual information, but are also effective ways to mobilise domain and policy expertise.

The Commission should be mindful of the need to apply different approaches and metrics to tackle different instruments and programmes, and to estimate the wide range of expected effects. Harmonisation is welcome, but flexibility is key.

'Survey fatigue' is increasingly perceived as a problem, generating a pressure for evaluation tools that are quantitative yet 'non-invasive' in the sense of not requiring any activity by those being evaluated. Ensuring that surveys are creative, as short as possible and high in quality is part of the needed response. Another needed component is for the EC to make it clear that beneficiaries' active participation in evaluation is expected or even required. Alternatives to surveys do not normally do the same job. For example, public consultations are uncontrolled, are rarely able to ask the right questions and involve massive response bias (those who want to make a point respond, others do not). Other non-invasive techniques are unable to address many of the needed evaluation questions.

The evaluation toolbox is being expanded with new and exciting – sometimes fashionable – approaches, including text mining, semantic analysis, social network analysis and Altmetrics. We have also seen a higher demand for methods usually used outside research and innovation studies, such as counterfactual methods. Any developments that push harder to find better ways to estimate and explain the value of FPs and their level of additionality are potentially interesting. Some of these techniques are in their infancy. When they are used, this should be based on a good understanding of their strengths and limitations, including their robustness and the degree of clarity available about what they actually measure. Establishing their usefulness depends upon deliberate experimentation and evaluation rather than "trying it and seeing what happens".

It is important to experiment but we must also learn. The emergence of new techniques provides new ways potentially to understand and measure the wide-ranging effects of the FP. However, those techniques should be brought in with an explicit consideration of their strengths and limitations, as well as their alignment with an overall robust evaluation framework.

Improvement of ● ● ● the collection of monitoring data

In recent years the EC has made great strides in collecting and standardising project monitoring data and in generating more access to them. Further efforts are necessary in order to make them more appropriate, relevant and openly available to evaluators and others.

While there is value in standardising a small number of indicators (such as budget data and counts of scientific publications) that are relevant in most parts of the FP, the bulk of the effort needs to be devoted to designing and collecting a small number of SMART (whether quantitative or qualitative) and verifiable indicators that are tightly connected to a well-defined intervention logic. The nature of relevant indicators will vary considerably, for example among the three main 'sectors' of Horizon 2020 (Excellent Science, Industrial Leadership, Societal Challenges), among which intervention logics and expected impact are very different. Large numbers of semi-relevant indicators are useful neither for monitoring nor for evaluation. Unreliable indicators such as participants' guesses about job creation are equally unhelpful.

The Commission should demand more precision, rigour and selectiveness in the choice of indicator used for projects and programme. This would improve programme design as well as monitoring and evaluation.

The Commission has made significant progress in collecting and centralising indicators for FP monitoring and evaluation. Particularly noteworthy is the development of the RESPIR-SESAM database. These efforts appear to be little understood in the research and evaluation communities. There is scope for improvement in the quality, consistency and completeness of data collected. A particular concern is the difficulty of matching different data about the same organisations. Unique identifiers such as the ORCID researcher identifier would help here. Nonetheless, this work presents an important opportunity for evaluation work.

There is an increasing number of FP instruments that aim to align European and Member State R&I policies, in particular the various joint programming and public-to-public partnership initiatives. This asks for more collaboration to assess outputs and impacts not only at European level but also at the Member State level. Jointly defining appropriate indicators and monitoring data with Member States for these instruments would allow for a better understanding of impacts across the European Research and Innovation Area.

The Commission should extend its efforts to improve the quantity and quality of monitoring data, to communicate about their existence and to make them as open as possible.

● ● ● Reinforce evaluation governance

To grasp the opportunity to be at the forefront of evaluation approaches in research and innovation policy the European Commission should boost its own evaluation competences and improve its governance structure.

The size and breadth of the R&I-framework programme naturally generates a large number of evaluations that at meta-level are meant to provide insights in the results and impact of the FP. In FP7, approximately 140 evaluations and assessments an estimated budget of more than €20m⁴ have been carried out. While H2020 has only been running for three years, there have already been a large number of studies and assessments. The interim evaluation alone consists of more than 20 individual programme evaluations and numerous horizontal studies and assessments.

The wide range of individual evaluations requires careful planning and coordination from the start. Not only should sufficient financial and human resources be available to support and coordinate these studies, but the evaluations should also be prepared and organised in timely manner. This includes preparing Terms of Reference and commissioning the study to an appropriate research organisation or composing a relevant expert group, allowing sufficient time for the actual study to take place, and developing in advance a strategy for using the results of the evaluation.

While recognising the complexity of managing evaluation across the FP, we nonetheless urge the Commission to ensure that the planning cycle is observed and that evaluations are launched early enough to allow them to be undertaken efficiently and effectively.

Evaluations need to be credible and legitimate as well as being objective and independent. These things stem from both political needs and from the practical need to ensure that evaluation judgements are not clouded by self-interest or personal commitment to the aims and objectives of the interventions being evaluated.

We emphasise the importance of involving independent expertise so as to avoid any biases in the design or implementation of the study, ensure a critical objective assessment of the topic at hand, enhance the credibility of the study, and most importantly generate the learning effects that evaluations ultimately aim to achieve.

Many different DGs, thematic units, and Executive Agencies are involved in commissioning and managing the range of evaluation studies within the FP. Evaluations of individual programmes or horizontal issues are typically designed, tendered, and management by the units or Executive Agencies that are responsible for them. They play an important role in ensuring that the terms of reference (in particular the evaluation questions) are interpreted correctly, the required depth and breadth of the study is achieved, questions from the evaluation team are answered, and the overall quality of evaluation studies is high. The experience from FP7 as well as the on-going interim evaluation of H2020 shows that the form and quality of individual programme evaluations is variable.⁵ The extent to which evaluation expertise and planning is spread across relevant parts of the EC is a crucial component of an effective monitoring and evaluation framework, in order to ensure that they are sufficiently informed and prepared to provide the necessary guidance to the evaluation team.

While centralised evaluation expertise is important, a degree of evaluation and evaluation management capability is needed across all units whose activities are liable to evaluation. The Commission should make efforts further to train and coordinate those involved.

We acknowledge and welcome the advisory role that centralised evaluation units within individual DGs play. However, in practice, evaluation approaches differ among DGs. This leads to inconsistency in the evaluation of the FP as a whole, complicates evaluation of cross-DG programmes and impedes the generation and sharing of common data. In order further to strengthen the coherence of monitoring and evaluation activities within the framework programme, it would be worth considering setting up a more structured approach to facilitate the exchange of information across DGs and Executive Agencies involved in the framework programme.

The Commission should investigate the opportunities to set up a common or central evaluation group, coordinating monitoring and evaluation across the FP as a whole.



End notes

1 EU Expert Group, (2015), Commitment and Coherence, Essential ingredients for success in science and innovation, Ex-Post Evaluation of the 7th EU Framework Programme (2007-2013), Brussels, page 11.

2 European Commission, (2015), Better Regulation Guidelines Commission Staff Working Document, Strasbourg, COM (2015), 215 final.

3 Erik Arnold, Bea Mahieu, James Stroyan, David Campbell, Malin Carlberg, Flora Giaracca, Andrej Horvath, Zsuzsa Jávorka, Paula Knee, Ingeborg Meijer, Sabeen Sidiqi and Caroline Wagner, Understanding the Long-Term Impacts of the Framework Programme, Brussels: EC, 2011

4 EU Expert Group, (2015), Commitment and Coherence, Essential ingredients for success in science and innovation, Ex-Post Evaluation of the 7th EU Framework Programme (2007-2013), Brussels.

5 EU Expert Group, (2015), Commitment and Coherence, Essential ingredients for success in science and innovation, Ex-Post Evaluation of the 7th EU Framework Programme (2007-2013), Brussels.

