SMART INNOVATION:
A Practical Guide to Evaluating Innovation Programmes

Study coordinators:
Louis Lengrand & Associés

Study partners:
PREST
University of Manchester
ANRT (Association nationale de la recherche technique)
Reidev Ltd

January 2006
SMART INNOVATION – Supporting the Monitoring and Evaluation of Innovation Programmes

This Guide was compiled on behalf of the project team and under the inspiration of ideas developed in HLWG meetings, by the PREST team. The main contributors are Ian Miles and Paul Cunningham, with inputs from Deborah Cox, Christina Crowe and Khaleel Malik. Editing support was provided by Marco Jaso.

© ECSC-EC-EAEC Brussels-Luxembourg, 2006

"Neither the European Commission, nor any person acting on behalf of the Commission, is responsible for any use which might be made of the information in this report.

The views expressed in this report are those of the authors and do not necessarily reflect the policies of the European Commission".
EXECUTIVE SUMMARY

This Guide is in effect a Handbook for those concerned with commissioning, implementing, or using evaluation and monitoring in the context of innovation programmes. As such, it comprises a large collection of Question-and-Answer formulations dealing with issues that arise when bringing evaluation approaches to bear here. This means that a conventional Executive Summary is not really feasible, and thus here we will focus on explicating the main structure of the Guide and the critical elements of the philosophy behind it.

The structure of the Guide is straightforward. First the nature of innovation programme evaluation is introduced, and the reasons for undertaking it set out. Arguments against such evaluation are addressed. Different approaches to evaluation are set out and the scoping of evaluation studies examined. The range of methods that can be brought to bear are briefly set out, together with a discussion of the issues that arise in terms of impact assessment. Finally the Guide concludes with a discussion of the reporting, use, and embedding of innovation programme evaluation.

The main philosophy that underlies the Guide was crystallised in the course of the current study, and draws very much on inputs we received from interviewees and our High Level Working Group. This can be summarised in terms of a number of key points:

♦ The culture of evaluation is developed to a very uneven extent around Europe, and while this suggests that there are countries and regions that are underutilising the approach, at least it means that there is good practice we can draw upon.
♦ In the more advanced cultures, evaluation has moved on beyond being a simple auditing of performance, and is becoming an integral part of a learning-based approach to policymaking and programme formation.
♦ Evaluations can cumulatively provide increasing insight into the operation of the innovation systems within which innovation programmes are seeking to operate.
♦ Innovation programme evaluation can learn a great deal from experience in research programme evaluation, but this area of evaluation poses challenges of its own that need to be addressed.
♦ There is no “magic bullet” in evaluation: no single method that can answer all of the main questions of an evaluation study, and be applied in all types of study. Typically, evaluations will need to use a mixture of methods chosen to fit the needs of the particular study.
♦ Effective use of evaluation requires informed users, who appreciate the necessary limitations of any individual evaluation study.

The informed users mentioned in the last bullet point will value the contributions of evaluation to the policy process, and are likely to voluntarily serve as ambassadors for innovation programme evaluation. It is hoped that this Guide will be helpful to such policymakers as well as for actual and potential evaluation practitioners.
Table of Contents

Foreword ........................................................................................................... 9

Acknowledgements ....................................................................................... 11

Definitions ...................................................................................................... 13

ACRONYMS .................................................................................................... 17

Chapter 1  INTRODUCTION........................................................................... 19

1.1 Why has this Guide been produced? .................................................. 19
1.2 Who is this Guide for? ........................................................................... 20

Chapter 2  WHAT IS THE EVALUATION OF INNOVATION
PROGRAMMES? ............................................................................................ 21

2.1 Why is Innovation a critical issue now? ........................................... 21
2.2 Why are Innovation Programmes a critical issue now? .................. 22
2.3 Why are Innovation Policies so diverse? ......................................... 23
2.4 How does Innovation Policy differ from Research Policy? ........... 25
2.5 What are the origins of Innovation Programme Evaluation? ....... 26
2.6 Does Innovation Programme Evaluation draw on Research Evaluation? ... 28
2.7 Why is Innovation Programme Evaluation a critical issue now? ...... 29
2.8 Just what does Evaluation mean? ....................................................... 31
2.9 Can we really hope to Evaluate something as complex as Innovation? ...... 32

Chapter 3 WHY EVALUATE INNOVATION PROGRAMMES? .................... 34

Arguments Against Evaluation .................................................................... 34

3.1 Why might people oppose evaluation? ........................................... 34
3.2 What are the detailed arguments against evaluation? .................. 35
3.3 How can the arguments against evaluation be countered? .......... 36
3.4 How can the arguments about the difficulty of evaluation be countered? ... 37
3.5 How can the arguments that policymakers do not need evaluation be countered? ................................................................................. 38
3.6 How can the arguments that evaluation is a misuse of essential resources be countered? ................................................................. 39
3.7 How can the arguments about the (lack of) necessity of evaluation be countered? ......................................................................................... 40

Arguments for evaluation ............................................................................. 41

3.8 How can we argue for evaluation? ....................................................... 41
3.9 What are the arguments around ‘Assessing value-for-money’? ........ 42
3.10 What are the implications of Evaluation goings beyond Auditing and Monitoring for arguments about 'Assessing value-for-money'? ..........43
3.11 What are the arguments about 'Improving programme management'? ....45
3.12 What is the role of Evaluation in 'Improving programme management'? ....46
3.13 What are the arguments about 'Improving design of programmes'? ...........48
3.14 What is the role of Setting Objectives in 'Improving design of programmes'? ........................................49
3.15 What is the role of Evaluation in 'Improving design of programmes'? ....50
3.16 How can we argue about 'Enhancing policy design'? .................................53
3.17 What other arguments are there for evaluation? .........................................54
3.18 What, then is the overall case for evaluation? .............................................56

Chapter 4 EVALUATION APPROACHES ....................................................58
4.1 How does Evaluation differ from Monitoring? ..............................................58
4.2 How does Evaluation differ from Auditing? ..................................................59
4.3 How does Evaluation differ from Analysis of Success Stories? .....................60
4.4 What are the Main Types of Innovation Programme Evaluation? ...............61
4.5 What is Ex Ante Evaluation? .....................................................................63
4.6 How can Ex Ante Evaluation Work? ..........................................................64
4.7 What are Interim, Periodic and Real-Time Evaluations? .............................65
4.8 What is Ex Post Evaluation? .....................................................................67
4.9 What does it mean for Evaluation to be an integrated part of policy making? 68
4.10 How does Innovation Programme Evaluation differ from Research Programme Evaluation? ........................................69
4.11 How far can we use the methods and messages from Research Programme Evaluation in Innovation Programme Evaluation? .................71

Chapter 5 SCOPING THE EVALUATION OF INNOVATION PROGRAMMES .................................................................72
5.1 What is the evaluation for? ........................................................................72
5.2 What is the scope of an evaluation? ............................................................74
5.3 What should be done before conducting an evaluation? .............................76
5.4 Why is a Programme Overview Important? ..............................................77
5.5 What is a ROAME statement? ..................................................................79
5.6 When should ROAME be used? ...............................................................80
5.7 What is the “Business Case” Approach? ....................................................81
5.8 What is the role of the Programme Overview? ...........................................82
5.9 What is the Balanced Scorecard Approach? .............................................83
5.10 How can we make statements of objectives more useful for evaluation and monitoring? ............................................................85
5.11 What criteria should be used for the evaluation? ....................................87
5.12 What about the financial costs of evaluation? .........................................88
5.13 What should the timing of an Ex Ante evaluation be? ...............................89
5.14 What should the timing of a Real-Time Evaluation be? .............................90
5.15 What should the timing of Periodic Evaluations be? ...................................91
5.16 What should the timing of an Ex Post evaluation be? .................................92
5.17 How can the objectives of individual evaluations be specified? ..................94
5.18 What types of indicators should be selected? .............................................95
5.19 What are the advantages of a quantitative focus? .....................................96
5.20 What are the advantages and disadvantages of a quantitative focus? ......97

Chapter 6 CHOOSING AND USING METHODOLOGIES FOR EVALUATION

6.1 Why is Methodology Important? ..............................................................100
6.2 How does the Evaluation Mission relate to Methodology? ....................102
6.3 What is the significance of Operational Definitions? ..............................103
6.4 Why is there no single Methodology for evaluation? ............................104
6.5 What are the Main Methods used in Evaluation? ....................................105
6.6 What sorts of Measurement are we looking at in Innovation Programme Evaluation? .................................................................107
6.7 What are the sorts of secondary data that may be accessed? ....................109
6.8 What are the Methods for generating data via primary research? ..........110
6.9 What is the Use of Interview-based methods for generating data in Evaluations? .................................................................113
6.10 How and when should Focus Groups and Workshops be used for generating data? .................................................................115
6.11 How and when should Observational Methods be used for generating data? 117
6.12 How and when should Surveys be used for generating data? ....................118
6.13 What are the Methods for structuring and exploring interventions in the course of generating data? .................................................................120
6.14 What are the Methods for analysing data? ............................................122
6.15 What are the advantages of combining qualitative and quantitative approaches? .................................................................124
6.16 What is Cost-Benefit Analysis? ..............................................................128
6.17 How can Cost-Benefit Analysis be Used? .............................................130
6.18 Do the Methods we use vary according to the type of Evaluation? .......132
6.19 What Methods are Useful in Ex Ante Evaluation? .................................133
6.20 What Methods are Useful in Monitoring? .............................................135
6.21 What Methods are Useful in Real-Time and Interim Evaluations? .......137
6.22 What Methods are Useful in Ex Post Evaluation? ..................................138
6.23 What is the Case Study approach? ..........................................................141
6.24 What sort of approaches are Participatory Monitoring and Evaluation? ....143
6.25 Can we evaluate multiple programmes? The role of Cross cutting evaluations .................................................................146

Chapter 7 EVALUATING IMPACTS ............................................................148

7.1 What is Impact Assessment? .................................................................148
SMART INNOVATION – Supporting the Monitoring and Evaluation of Innovation Programmes

7.2 What Challenges does Impact Assessment face in terms of Problems of Complexity and Causality? ................................................................. 151
7.3 What Challenges does Impact Assessment face in terms of Time Lags? .... 153
7.4 What Challenges does Impact Assessment face in terms of Indirect Effects? 154
7.5 What Challenges does Impact Assessment face in terms of Multiple and Complex Objectives? ................................................................. 155
7.6 What sorts of impacts can be assessed, to what extent? ......................... 156
7.7 What is Additionality? ........................................................................... 159
7.8 What is Behavioural Additionality? ....................................................... 160
7.9 How does evaluation provide an insight to a systemic process? .............. 162
7.10 What have we learned about innovation programme evaluation and the culture of evaluation? ................................................................. 164

Chapter 8 REPORTING ON EVALUATION RESULTS ......................... 166
8.1 Why is it important to think about Reporting? ........................................ 166
8.2 How should reporting requirements be decided? ................................ 169
8.3 How can we best present the results of an Evaluation to a particular audience? ................................................................. 170
8.4 What does a report contain? ................................................................. 171
8.5 How can results be presented? ............................................................. 173
8.6 What should be done by way of consultation and feedback? ................. 174
8.7 What should be done about confidentiality? ........................................ 175

Chapter 9 MAKING THE CASE FOR EVALUATION ......................... 176
9.1 How can we deal with Arguments against Evaluation? ....................... 176
9.2 Who needs to be convinced? (1) champions ........................................ 177
9.3 Who needs to be convinced? (2) stakeholders .................................... 178
9.4 What are major challenges that have to be overcome to win Support for the Use of Evaluations? ................................................................. 179
9.5 How can we address problems of Timing? ........................................... 180
9.6 How can we address problems of Resources? .................................... 182
9.7 How can we address problems associated with Inappropriate Tools and Instruments? ................................................................. 183
9.8 How can we address problems of Quality Control? ............................ 184
9.9 Are there Standards for Evaluators to Follow? .................................... 185
9.10 How can we address problems of Independence? ............................... 189
9.11 What problems may arise in terms of unrealistic expectations of evaluation? ................................................................. 191
9.12 How can we contribute to building an evaluation culture? ................. 192
9.13 Does it make sense to evaluate Programmes in isolation, or should an Evaluation Function be institutionalised? ........................................ 193
9.14 How can we affect how far the results of Innovation Programme Evaluations are applied? ................................................................. 194

BIBLIOGRAPHY ...................................................................................... 196
Foreword

The SMEIP project – SMEIP stands for Supporting the Monitoring and Evaluation of Innovation Programmes – has been funded by DG Enterprise. The project has included case studies of evaluation systems, interviews and an online survey of evaluation practitioners and stakeholders, and, crucially, a series of High Level Working Group meetings. A well-attended conference was also held in Brussels in early summer 2005. These have all proved vital resources in forging the idea for this Guide – creating a usable resource in addition to our conventional project reports.

One of the main goals of the project was to identify opportunities for the European Commission to encourage future evaluation practice in the innovation programme field. While the project reaches numerous conclusions as to strategies for enhancing the culture of evaluation, one major commitment was to deliver a plan for a pilot initiative that would promote this goal.

This Guide, Smart Innovation, is intended to be of wide use, but its immediate application will, hopefully, be in the pilot initiative (should this be undertaken). In this context, the Guide should be seen as a dynamic entity, something that will be honed and improved through experience gained from applying it. More prosaically and conventionally, we could say that this version of the Guide is a first draft, and that we are hoping for feedback on its structure and contents which will allow for improved and augmented editions in the future.

One obvious direction for further development of the Guide is to configure it as a hypertext – an electronic tool to be accessed via the Internet or on a CD-ROM. Such a version could be linked to the other materials of the SMEIP project as they are put online, and to the rich corpus of relevant materials generated elsewhere. (One of the recommendations of SMEIP is that more effort being put into documenting and disseminating information on the practice and use of innovation programme evaluation and evaluation systems more generally). In the short time of the present project, we have not been able to render this document as a hypertext, but we hope that the structure will be sufficiently clear and user-friendly for users to apply it effectively in paper or PDF format.
Acknowledgements

First, thanks to our project officer, Mr Alberto LICCIARDELLO (Innovation Unit, DG Enterprise).

Especial thanks to our High Level Working Group Members, who helped define the tasks of this Guide and provided us with insights and much valuable material:

♦ Mrs Agnes ARCIER, Head of Unit, Innovation and Competitiveness Policies, Ministry of Economy, Finance and Industry – France
♦ Mrs Ullrike BLANKENFELD, Unit VIII D 1, Federal Ministry of Economics and Labour – Germany
♦ Mr Lennart NORGREN, Analyst, Vinnova, representing Mr Kjell-Håkan NÄRFELT, Board Member and Advisor, Vinnova – Sweden
♦ Mr Arjan WOLTERS, Policy maker/evaluation expert at the Ministry of Economic Affairs, representing Mr Theo ROELANDT, Director of the Department of Strategy, Research and International cooperation of DG Innovation, Ministry of Economic Affairs - Netherlands
♦ Mr Paolo GUARNIERI, Operative Unit, European Regions Knowledge Innovation Network, and Simone SORBI, Director of Innovation and Research Policy Office, Toscana Region – Italy
♦ Mr Janis STABULNIEKS, Managing Director of the Latvian Technological Center – Latvia
♦ Mr Peter STANOVNIK, Vice-Chairman, National Council for Technology – Slovenia
♦ Mr Neil MACCALLUM, Head of Appraisal & Evaluation, Scottish Enterprise – United Kingdom, representing Charlie WOODS, Senior Director, Knowledge Management, Scottish Enterprise – United Kingdom

The other Study Team Members contributed material and motivation: thanks to Mr Louis LENGRAND, coordinator and Ms Julie KOELTZ, (Louis Lengrand et Associés); Mr Alain QUEVREUX, and Mr Jean-Marc LEVY, (Association Nationale de la Recherche Technique), and Mr Reinhold ENQVIST, (REIDEV Ltd).

Finally, the Interviewees, Survey respondents and Case Study participants are thanked for their helpful inputs and information. They are too numerous to name individually, but we thank you all sincerely.
Definitions

**Innovation** in industry is a matter of doing new things, or finding new ways of doing familiar things. Much of the discussion of innovation revolves around **product innovation** (the creation of new or improved goods and services) and **process innovation** (new ways of producing goods and services). However, there are also innovations in terms of **interfaces** between organisations and between suppliers and users of products (marketing, e-commerce, new systems for delivering goods and services, after sales services, interactions between suppliers and users concerning product design and specification, etc.) **Organisational** innovations are sometimes differentiated from **technological** ones (though they often go hand in hand). Another important distinction is between **incremental** innovations (minor modifications to products and processes) and more **radical** innovations (major changes in how processes are organised and how products work). Incremental innovations often emerge from the experience of making and delivering products; radical innovations often require Research and Development or similar efforts to bring new knowledge to bear on the product or process. An idea or project is not an innovation unless it is applied in processes put onto the market, or used in the public sector. (In the business world, it is common to restrict use of the term to **successful** applications only: but we believe there can be much to learn from innovations that are unsuccessful in terms of attaining the expected markets, or even in terms of failing to meet technical specifications.)

**Innovation Programmes (IPs)** are measures, schemes, initiatives, etc. Funded by (any level of) government, aimed at the promotion, support or stimulation of innovation and innovation-related activities. They may operate either directly, through the provision of funding, information or other support; or indirectly, through facilitation of the innovation process (i.e. via fiscal or regulatory reform). Note that some Innovation Programmes may have innovation as a secondary objective, or as a means to an end such as greater energy efficiency, or regional development.

**Programme Overviews** encapsulate in a few pages or in a relatively simple diagram what a programme is, and what it is trying to achieve. They include approaches such as Programme Models or Logical Frameworks. One type of Programme Overview is known as the ROAME statement. Prepared at the initial planning stage of the Programme, such a statement will set out the Programme’s:

- **Rationale** – what is its overall goal, what policy aims is it trying to support.
- **Objectives** – what are the specific achievements the Programme is intended to be making in support of this Rationale?
- **Appraisal** – what are the activities that will realise these achievements to be selected?
Monitoring – what are the means for examining the progress of the different activities, in terms of meeting milestones and performing the allocated tasks.

Evaluation – what approach will be adopted to examine the extent to which the Programme has achieved its objectives effectively and efficiently?

Whether put within a ROAME framework\(^1\) or some other layout, a Programme Overview that is established at the outset of an Innovation Programme should make the Programme easier to manage and, ultimately, to evaluate. It may well also be a useful way of communicating about the Programme to financial and other political authorities, and to other stakeholders.

**Evaluation** in this context refers to **Programme Evaluation**. This is the systematic assessment of a programme to determine (1) how far it is meeting objectives (and, perhaps, achieving other effects), (2) whether it is meeting these objectives efficiently and effectively), and (3) how the management structures (and other factors) have shaped these results. Evaluation may be ex ante, monitoring, ex post, etc.; and it may focus on a single project or programme, or extend more broadly to encompass the policy framework within which these are located.

**Evaluability** refers to how easy it is to evaluate something – this will depend, for instance, on how far there are clearly defined objectives against which achievements can be assessed, how widespread and diffuse these effects may be, over what time period they may be expected to emerge, and so on. It is typically easier to evaluate the immediate consequences of a programme than its longer-term outcomes – though the latter are probably what motivates the programme in the first place, particularly in the case of innovation programmes.

**Innovation Programme Evaluation (IPE)** is the evaluation of Innovation Programmes – each of these component terms has been defined above. While many of the tools and techniques that are used here derive from those employed in the evaluation of R&D programmes, the aims of Innovation Programmes are typically wider than those of R&D programmes, and there are fewer standard indicators that can be used for evaluation purposes. For instance, R&D programmes can be assessed in terms of contributions to knowledge like publications and patents; but IPs require that their impacts on, for example, business practices and performance, are assessed.

**A culture of evaluation** is a term that is used to refer to how far evaluation practices are embedded into an institution’s policy and programme design and implementation. In an advanced Evaluation Culture, evaluation is treated as a natural and inherent feature of programmes, and planned for at the outset of programmes. It is not just something that is simply added on at the end of the exercise – though to conduct any evaluation at all is one step up from an

\(^1\) More detail on ROAME and analogous approaches is provided below in Chapter 5.
institutional framework in which there is simply no evaluation at all. In more advanced Evaluation Cultures, furthermore, evaluation is not just a way of assessing the performance of a particular IP. It has become a tool for informing the design of IPs, and indeed informing innovation policy, more generally.

**Additionality** is an important element in considering the effects of a Programme. Additionality is the change due to the activity, as compared to what would have happened had the activity not been undertaken at all. More discussion of additionality is provided in the text below (Chapter 7).

**The Linear Model of Innovation** is based on the notion that predominantly, innovations emerge from the elaboration of increasingly practical applications of new fundamental knowledge. Innovation is typically triggered by discoveries made in research laboratories that are found to have potential use in creation of new products and processes. Stimulating innovation is then, largely a matter of pump-priming R&D: the new knowledge will be converted into innovation by entrepreneurs. Many innovation studies – and evaluations of R&D programmes – have cast doubt on this account, and a number of more complex models have been proposed. These incorporate, for example, all sorts of feedback loops, and the likelihood that innovation can be initiated at any point in what was earlier seen as a sequence or chain of activities – even by users. But while the linear mode is habitually criticised in the research literature, and has even been rejected in official documents, it is still implicit in a great deal of policymaking.
### ACRONYMS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>Advanced Technologies Programme</td>
</tr>
<tr>
<td>DTI</td>
<td>Department of Trade and Industry (UK)</td>
</tr>
<tr>
<td>EC</td>
<td>European Commission (also known as CEC)</td>
</tr>
<tr>
<td>EPEC</td>
<td>European Policy Evaluation Consortium</td>
</tr>
<tr>
<td>ePUB</td>
<td>An EC-funded network on the knowledge accumulated in the area of Socio-Economic evaluation of public RTD policies</td>
</tr>
<tr>
<td>ESPRIT</td>
<td>The European Commission’s information Technologies programme, spanning industrial R&amp;D projects and technology take-up measures</td>
</tr>
<tr>
<td>EUREKA</td>
<td>A European network for market-oriented, industrial research and development</td>
</tr>
<tr>
<td>HEI</td>
<td>Higher Education Institution (e.g. a University)</td>
</tr>
<tr>
<td>IKED</td>
<td>International Organisation for Knowledge Economy and Enterprise Development</td>
</tr>
<tr>
<td>IP</td>
<td>Innovation Programme</td>
</tr>
<tr>
<td>IPE</td>
<td>Innovation Programme Evaluation</td>
</tr>
<tr>
<td>IT</td>
<td>New Information Technology (also sometimes referred to as ICT, IST)</td>
</tr>
<tr>
<td>NSF</td>
<td>National Science Foundation (US)</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>Research and Development (also sometimes referred to as RTD)</td>
</tr>
<tr>
<td>ROAME</td>
<td>Rationale, Objectives, Appraisal, Monitoring, Evaluation</td>
</tr>
<tr>
<td>SMEIP</td>
<td>Supporting the Monitoring and Evaluation of Innovation Programmes - the name of the project from which this Guide derives.</td>
</tr>
<tr>
<td>TEKES</td>
<td>The National Technology Agency of Finland</td>
</tr>
<tr>
<td>VINNOVA</td>
<td>The Swedish Agency for Innovation Systems</td>
</tr>
</tbody>
</table>
Chapter 1 INTRODUCTION

1.1 Why has this Guide been produced?

Over recent years innovation policy has risen to the fore. It is now widely recognised that competitiveness and other socioeconomic goals are heavily dependent on industrial innovation. Government initiatives to support innovation have proliferated. Another thing that is now widely recognised is that such initiatives go well beyond Research and Development (R&D) support. Innovation Programmes take many forms, attempting to influence various parts of (what have become known as) innovation systems, with a wide range of objectives and instruments.

Evaluation, at best, is a learning process helping to design better programmes in the future, and as a consequence save taxpayers money while achieving important socioeconomic objectives. The need to evaluate Innovation Programmes is becoming widely recognised, though there are still many environments in which this is not the case. Even more prevalent is a very narrow view of what evaluation could be and what it should contribute. There are relatively few environments where a “culture of evaluation” has been well-established.

The aim of this Guide is to help build more of a culture of evaluation of Innovation Programmes. It discusses the reasons for undertaking such evaluations, and how the arguments that may be mounted against them can be countered. It sets out the different functions which evaluation may be used for, and suggests how these may underpin a culture of evaluation. It sets out the methods that can be used in evaluations, pointing out that Innovation Programme evaluation is inevitably a complex affair. (R&D Programmes are relatively simple to evaluate, since the key outcomes are the accumulation of knowledge and scientific prestige. The sorts of impact on innovative performance that are aimed at in Innovation Programmes are multifaceted and relatively hard to trace.) It discusses the design and management of evaluations, and how they can be used effectively in forming programmes and policies.

This Guide seeks to crystallise available knowledge about the issues confronting the evaluation of Innovation Programmes, drawing out the practical implications that follow from this knowledge. It draws on the results of literature reviews, a survey, interviews and case studies conducted in the course of the present study, and the experience of the project team in evaluation practice. By crystallising practical knowledge gained over many evaluation studies and activities, the Guide aims to lower some of the barriers to evaluation exercises, and to contribute to building more sophisticated cultures of evaluation.

Last but not least, this Guide would never have seen daylight unless there is a widespread view among experts in the EC and outside it that more evaluation in the field of innovation is needed to make Europe competitive enough to guarantee its inhabitants welfare in years to come.
1.2 Who is this Guide for?

“Smart Innovation” is intended to be a “Practical Guide” that will serve as a reference point for those interested in conducting, using and/or setting up Innovation Programme evaluations. Readers might be concerned with organising an evaluation of a particular programme, or with putting in place an evaluation system in a particular organisation. You could be concerned with improving the quality or the use of evaluations that your institution undertakes, or with making use of the results of a series of evaluations undertaken by third parties. You could be a policymaker for innovation policies, a programme manager, an evaluator (or would-be evaluator) yourself. You could be based at national, regional, or even transnational levels.

We have aimed to make the Guide appealing and useful for such a variety of different readers. Thus, it has been designed so that it can be used in several different ways. The modular, question-and-answer format, should make it easy for you to dip into different parts of the Guide. It can be read sequentially, from beginning to end, or you can turn to those sections and questions that interest you most. Cross-references are included throughout to indicate where other chapters and sections discuss complementary aspects of given issues.

If you do not have much knowledge about evaluation and how it can be used, the Guide introduces and explains, illustrates and exemplifies, evaluation practice and the case for evaluation (Chapters 2 and 3, especially).

Chapter 4 introduces some of the different approaches to evaluation that can be adopted with respect to Innovation Programmes.

If your interests are more immediately instrumental, and you wish to use the Guide to spell out the major issues to be confronted in programme design and management, you might consider focusing initially on Chapter 5. The methods used in evaluation studies are discussed in Chapters 6 and 7, while how the reporting and use of evaluation results, so that their contributions to policy can be maximised, is taken up in Chapters 8 and 9.
Chapter 2 WHAT IS THE EVALUATION OF INNOVATION PROGRAMMES?

2.1 Why is Innovation a critical issue now?

Innovation has risen to the fore in thinking about the future prosperity and quality of life of countries and regions across Europe and, indeed, world-wide. It is now widely accepted that economic performance is not just a matter of access to natural resources and large markets, or even of having a skilled labour force on tap. These things are important, but much of the achievements that have been made in terms of economic growth and social welfare are related to innovation. New processes have made it possible to create mass markets for products that were previously the preserve of elites; new products have allowed people to achieve higher living standards. Innovative companies are more likely to grow and prosper: countries and regions with more innovative environments are more dynamic and successful. These views are now so widely accepted that it may come as a surprise to realise that only a few decades ago they were regarded as rather peculiar ones. It has taken many economic and industrial studies, and much practical experience, to change our worldviews.

Innovation is becoming even more prominent as an issue because of the advent of new technologies, around which new products and processes are proliferating. It may seem rather hackneyed to talk of the “Information Technology Revolution”, but the continuing increases in power and reductions in cost of the underlying technologies are so dramatic as to mean that we can expect to see immense changes continuing here in the coming decades. The new biotechnologies (derived from genomics and subsequent work) are beginning to find applications in health, agriculture, environmental services, and elsewhere – we are perhaps in much the same situation here as we were with information Technology in the 1970s. Nanotechnology – which covers a large range of very small things, from new materials through miniaturised electronics to nanomachines – is at an earlier stage of development, but applications from sun creams to extremely sharp cutting tools are already on the market. Successful economies will need to be able to make good use of these technologies, and where appropriate to be engaged in the supply and configuration of products based on them, and complementary goods and services supporting and enhancing their use.
2.2 Why are Innovation Programmes a critical issue now?

Governments of all political hues and regional dimensions have recognised the importance of innovation. Accordingly, they have sought to foster innovation in their territories.

As innovation has become a core goal of policy, so Innovation Programmes have become important tools for realising innovation policies. There is a problem here, however. Policymaking is traditionally compartmentalised into a set of ministries or government departments – education, science, industry, energy, environment, and so on. Innovation spans all these and more. There are rarely Ministers for Innovation in a senior position in governments. We do not often hear innovation policies and programmes being championed by leading politicians. The programmes that are most often explicitly known as Innovation Programmes will most often derive from ministries of industry or from regional governments (who are increasingly significant players in this area). But other sectors of government can also be active here. Innovation policy is often a rather complicated, and not always coherent, set of activities undertaken across several arms of government. Sometimes the promotion of innovation is a conscious aim of policies. Sometimes policies inadvertently affect innovation positively or negatively. One thing that evaluations are liable to uncover is that promising Innovation Programmes can be undermined by actions undertaken by other policy areas.
2.3 Why are Innovation Policies so diverse?

The simple answer is that innovation itself is a multifaceted thing. In different countries, there may be different inheritances of science and technology, different cultural assumptions about the role of the state, different attitudes to entrepreneurship, and so on. The international climate has changed; new technologies have emerged; and researchers and policymakers have learned more about innovation processes. It is not surprising that innovation policy has taken many different forms.

One traditional approach was simply to support “national champions”. The idea was that the biggest, highest-technology firms in one’s country just needed a little more encouragement (subsidies, preferential markets) to be able to keep up with the global frontier. This approach was largely discredited as the national champions often took the hand-outs, and repeatedly came back for more - while failing to adopt new practices. (Or, in many cases, if they did try really new things, they banked on standards and designs that were unable to break through into world markets.) It became politically difficult to support these big firms in overt ways, as trade and procurement regulations were adopted that reduced the scope for such favouritism.

The doctrine of national champions also suffered as the dominant political positions shifted toward recognition of the limits of state power in liberal democracies. It became accepted wisdom that the state was not in a position to second-guess the market as to which companies, which products would thrive. Support for national champions still continued – the big firms still possess political and economic clout. Many governments have wanted to avoid the slight to national pride that would supposedly be occasioned by the companies being merged into larger transnational corporations, or the lay-offs that would be threatened by withdrawal of lucrative contracts. But their support has become less overt and less legitimated by a claim that it is all about innovation.

The consensus around the limits of state power also dealt a blow to the idea of “picking winners” – the notion that governments were in a good position to decide which technologies were the ones that should be invested in across the economy. There might be reasons beyond economics to support particular technologies – for instance, to meet environmental targets, to provide wider access to key technologies through schools and libraries, and so on. But in general, the government was not in a position to say which technological promises were the ones that were most likely to satisfy market requirements. Policymakers had had their fingers burned often enough by entrepreneurs and visionaries, and the managers of national champions, assuring them that this or that product was the solution. The selection processes of the market were going to have to have their way, choosing among a range of products that were on offer. The most that governments could do was to identify the broad areas of technology where important opportunities were likely to develop.

Another factor concerns the system within which innovation is situated. Innovation does not reside in the hands of a limited group of key industrial
players but is a “team game” which draws on the activities of SMEs, universities, the public sector, competitors and collaborators, to name but a few. Innovation policies need to take account of these actors, the interfaces between them and the broader framework conditions which influence their behaviours.
2.4 How does Innovation Policy differ from Research Policy?

The disillusion with “picking winners” – national champions or specific products that the country should focus on – has been widely felt. When this point of view was articulated together with the “linear model” of innovation, it led to a focus on **Research and Development Policy** (sometimes just seen as Research Policy). The state, through this policy, could create national strength in the areas of scientific knowledge from which, the linear model suggested, commercial innovations would eventually flow. There have been several problems with this approach, most notably the tendency for high-quality science (if indeed this was achieved) to fail to connect well with local entrepreneurship. Sometimes the science would show no evident route to technological applications; this is not an argument against funding scientific discovery, but does mean that not all research should be justified as contributing to innovation goals, or be seen as part of innovation policy. Sometimes great scientific achievements – for instance in biotechnology – would be seen to have considerable technological pay-offs, but these would not interest local firms, and the knowledge would be acquired and applied by those in other countries. The linear model was thus failing in practical terms – as well as being subject to a ceaseless barrage of criticisms from innovation researchers.

A more sophisticated understanding of innovation processes began to be built into innovation policy, though we can still often find remnants of the linear model underpinning policy thinking. Instead of a linear flow of knowledge along a series of stages from science to commercialisation, innovation is seen as involving many feedback loops between these different “stages”. One goal for policy is to enhance communication across these points of action. Seeing innovation as a process taking place in a system, more recognition is also given to the diffusion of innovations, not just to their invention. Innovation policy has not just got to do with research, then, but also with such issues as commercialisation of inventions, with technology transfer, adoption and implementation, and so on. To the extent that it is still about research, innovation policy is focused on supporting R&D programmes that are much more closely linked to market and user requirements, and the actions needed to relate innovations to these. Innovation Programmes cover a much wider range of mechanisms and instruments than just support of R&D.
2.5 What are the origins of Innovation Programme Evaluation?

Innovation policy itself arose to prominence at a time when public budgets were coming under scrutiny. Public expenditure was seen as diminishing the funds available for reinvestment by the private sector, and as replacing consumer choice with government decisions. It needed to be justified, as achieving what it set out to do and not actually undermining the private economy. A major rationale for state intervention into economic affairs was that of “market failure”. Governments should only be active where markets were manifestly not operating in the way that economic orthodoxy believed they should.

For instance, governments might need to play a regulatory role where there were oligopolistic or oligopsonistic\(^2\) players at work in a market, who could use their power to fix prices, etc. They might need to provide information for economic agents when there was insufficient knowledge about market affairs (remember that the orthodox economic models began by assuming that parties have access to perfect information that they can process without cost).

Governments might need to act to supply various public goods where the returns to producers from the market alone would be insufficient to motivate adequate activity – these might be things like care for the impoverished elderly, economic stabilisation, or the setting of standards. One of the public goods most relevant to innovation is general and specialist education and training. (For example, it has been argued in some countries that businesses lack the incentive to provide advanced training for their staff because these trained staff may move on to benefit another firm that has not invested in training. Government is then seen to have a role in providing an adequate supply of skills.) Financing the production of knowledge which has no immediate economic value forms another public good highly relevant to innovation. Funding of basic research has long been seen, in large part, as a matter for government.

Research policy was subject to pressures on the public purse, and it was becoming apparent in the 1970s and ‘80s that priorities would need to be set among different lines of research. Governments could not continue to increase funds across all fields of study. But governments were not expert in research nor in making judgements as to whether funds were being used well, and decisions about promising areas of research to fund had traditionally been taken by scientists themselves. However, it was feared that scientists would resist setting priorities, or favour their own fields and colleagues. Thus it was necessary to bring in outside evidence to help assess whether funds were being well spent, and to set priorities. Evaluation was one of the major tools here, especially with respect to determining how efficiently and effectively research funds were used. (As we shall see, evaluation also feeds into informing priority-setting, though other tools – such as Foresight – have also become prevalent to underpin such decisions.)

---

\(^2\) The operation of markets can be seriously influenced when they are dominated by very few suppliers or customers. These can create pressures on prices, standards, and contractual arrangements that would not be viable in markets where there is more competition among numerous players.
Evaluation principles have been elaborated across many areas of public policy – a major stimulus here was the “Great Society” programmes of 1960s America, especially those programmes which aimed at boosting the performance of underachieving children from deprived communities. Evaluation tools for examining research policy took off in the early 1980s, as governments were starting up new R&D programmes to meet the challenges posed by new Information Technologies (IT).
2.6 Does Innovation Programme Evaluation draw on Research Evaluation?

The large-scale R&D programmes into IT that became prevalent in the 1980s dealt with topics that could not easily be categorised in the traditional dichotomy of pure and applied research. Progress in new IT involved fundamental new knowledge about a range of topics – the performance of semiconductors and optical fibres, the ways in which software could be structured and processed, the nature of user interfaces, and a host of new fields like mechatronics, chematronics, optronics, and the like. This new knowledge was often generated in response to requirements from IT suppliers and users, not just from the vagaries of scientific curiosity. It required basic scientific research, yet was driven by and found rapid application in practical purposes. The term *strategic research* became popular as a way of describing such work, and the parallel activities that have since been apparent in biotechnology and nanotechnology work.

Strategic research posed an interesting challenge for evaluation studies. On the one hand it was possible to evaluate the expenditure of funds in terms of its contribution to scientific knowledge. Were journal articles being published and cited, was the work having an effect on scientific communities? On the other hand, was it having the practical effects that were sought – was the work being taken up and applied by industry in commercial technologies? How significant were these? The question of research excellence was only one issue to be evaluated: funders also needed to take into account broader socio-economic consequences of the research.

Strategic research was being funded by governments more in order to achieve innovation objectives than those of pure science – though often these objectives were expressed in such terms as “riding the wave of the IT revolution” or “keeping pace with the opportunities of the new technology”. This reflects the fact that the effort focused on being innovative around one particular set of technologies – not being more innovative in general. Some Innovation Programmes are focused in this way, while others are much more generically oriented to innovativeness.

Innovation policy can include much research policy – though it makes sense to continue to fund basic research with no immediate economic rationale. Much of the experience and expertise, many of the tools and techniques, of Innovation Programme evaluation will have descended from research evaluation. But as it is necessary to address a wider range of programmes, and a wider range of impacts, so it has been necessary to develop new approaches to evaluation. The assessment of impacts on the behaviour of economic agents is a major focus of evaluation activity.
2.7 Why is Innovation Programme Evaluation a critical issue now?

Evaluation continues to be highly significant in the innovation policy arena. In part this reflects the increased prominence of evaluation across practically all areas of policy. This has something to do with the rise of the “knowledge-based society” – decisions are supposed to be based upon knowledge rather than prejudice or chance. “Evidence-based policy” has become a rallying call for modernising politicians. Governments are under pressure to justify their expenditure decisions to an informed media and populace. Their performance is benchmarked by a variety of public and private bodies, informing the investment decisions of global corporations, for example.

Evaluation is also significant because different parts of government are in effect competing for a limited share of public funds. They have to justify their planned expenditures to political authorities – not just to a prime minister or president, but also to the ministry of finance, and whatever “watchdog” organisations scrutinise public finance and accounts.

The figure below shows how public opinion perceives the importance of Innovation Programmes Evaluation in the framework of the formulation of innovation policy. A large majority of interviewed stakeholders within the survey conducted for the study agrees with the following statement: “Evaluations of innovation programmes are important in the formulation of innovation policy in my country “.

Figure 2.1 Evaluations of innovation programmes are important in the formulation of innovation policy in my country

Source: Supporting the Monitoring and Evaluation of Innovation Programmes survey data
But perhaps the major issue is that innovation programmes are often working in largely uncharted and poorly understood territory. While the importance of innovation is generally conceded, the nature of the innovation process is undergoing near-continual change. (For instance: sectors that were previously regarded as non-innovative – such as many services – are seen to be highly active (and new countries are joining in the process, too); the location of critical knowledge, and the sorts of networks that are formed to apply this in innovations, is evolving in new directions; new tools for managing innovation are becoming available.) Innovation policies span a huge range of actions and targets, and there is much learning needed about how to act effectively in diverse and changing conditions. Evaluations of those interventions that have been made is thus a vital tool.
2.8 Just what does Evaluation mean?

Evaluation is, at heart, a matter of seeing how well a policy or programme has achieved the objectives set for it. There is, however, a little more to it than this. Some authors (e.g. Polt 1998) differentiate between three aspects of evaluation – evaluation of effectiveness, of efficiency, and of efficacy. The features of evaluation, then, are:

♦ Evaluation can examine how **efficiently** the programme operated, exploring the management of the activities to see whether there was good communication of objectives and progress, avoidance of redundancy and ability to detect and address problems as they arose.

♦ Evaluation can examine the **effectiveness** of the programme, how far the objectives have been met. This is easier if the objectives have been set in quantitative or verifiable terms, and/or if there are related programmes against which it can be benchmarked. In assessing programme effectiveness, it is common to use **performance indicators**. Here the concern is, as has been well spelled out by the United Way of America (2002) with outcomes: "not how many worms the bird feeds its young, but how well the fledgling flies".

♦ Evaluation can go further and consider the **efficacy** of the programme: how relevant it is to the broad policy goals to which it was designed to contribute.

♦ Beyond these very pragmatic goals, evaluation can also involve **explanation**. It can cast light on how and why the programme achieved the degree of effect that was realised. It can illuminate issues of programme management, and, more profoundly, raise the understanding of the broader innovation system within which the programme has operated.

♦ Evaluation may be used in what may seem to be a simplistic manner, but can prove to be technically very demanding to operationalise, to attempt to examine the precise value for money that has been achieved (in terms of an estimate of the rate of return of the public expenditure).

♦ Evaluation can also be used to examine unintended consequences of the programme intervention: benefits and costs of the activities that were not expected by the programme designers, or not explicitly spelled out as being among the programme objectives.
2.9 Can we really hope to Evaluate something as complex as Innovation?

Many areas of policy have become subject to evaluation and to more routine monitoring of performance by a range of relevant indicators – school achievement, prisoner recidivism, hospital bed occupancy, levels of poverty and unemployment, and so on. Innovation policy is subject to the same pressures. But innovation is rather challenging in terms of performance indicators. How can we tell if we have increased the innovativeness of the economy?

Specific Innovation Programmes usually have more modest goals that are intended to contribute to this grand objective. For instance, programmes might set out to raise awareness of particular technological opportunities or of the potential benefits for firms and Universities to collaborate in developing new applications of emerging knowledge in a particular area. They might provide support for small firms active in an area to protect their intellectual property, for spin-offs to be generated from government laboratories, for consultancies to help industrial clusters use new approaches in their supply chains and other relationships. The programmes might be viewed as technology transfer policies, intellectual property policies, training policies, cluster policies and so on. But if – as in the examples mentioned – innovation is among the main goals, then they can also be seen to be Innovation Programmes. And as such, they can be evaluated in terms of the more modest goals they have – is the desired change effected, and does it appear to have the desired consequences?

Table 2.1 suggests ways in which the more modest goals of innovation programmes could be evaluated, and how these can be linked to longer-term, more ambitious, impacts.
<table>
<thead>
<tr>
<th>Objectives</th>
<th>Objects of Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Programme Aims</strong></td>
<td><strong>Immediate Impacts</strong></td>
</tr>
<tr>
<td>Increase awareness of a set of new technologies</td>
<td>Attendance at meetings, uptake of literature, hits on websites</td>
</tr>
<tr>
<td>Improve the skills base of a set of industries</td>
<td>Training sessions, staff exchanges</td>
</tr>
<tr>
<td>To increase HEI-industry links</td>
<td>Student placements, increased academic-industry dialogue</td>
</tr>
<tr>
<td>Stimulate the start-up of new-technology based companies</td>
<td>Finance and information for company start-up for would-be entrepreneurs</td>
</tr>
</tbody>
</table>

(There are many other possible objectives: the list above is illustrative.)
Chapter 3 WHY EVALUATE INNOVATION PROGRAMMES?

Arguments Against Evaluation

3.1 Why might people oppose evaluation?

For those who come from a well-developed culture of evaluation, it may seem strange, or even suspicious, that there are countries and government departments where evaluation is treated with some suspicion. Evaluators might concede that there are issues about how many evaluations are undertaken, how often, and at what cost – but the principle of evaluation is well developed and widely accepted in some cultures.

But this is not universally the case. Evaluation may be opposed on a number of grounds. Before we consider the intellectual case against evaluation, we need to admit that this can be used just a smokescreen. Restricting the use of evaluation can reduce the chances of exposure of inadequately designed or implemented programmes, of arbitrary policies that fail to learn from earlier successes and failures, of too close relationships between policymakers and those who stand to benefit from the programmes.

(Similarly, evaluation itself may be commissioned as part of a political point-scoring exercise, rather than as an effort to learn how to do things better. A new government or government appointee may seek an evaluation to demonstrate the incompetence or worse of its predecessor. This may be the opportunity to institute some serious evaluation practice, but it will be important in such instances to avoid letting the evaluation become just a political tool – evaluation is first and foremost a learning tool. It will need to be conducted systematically and as accurately as possible, and those who commission it must be prepared for the possibility that it will come up with results that do not serve their immediate political ends. If this is not the case, then we have another potential argument against evaluation – that it is liable to be merely attacking those in power in the past and/or legitimating those in power now.)
3.2 What are the detailed arguments against evaluation?

Some of the “common-sense” arguments against evaluation concern technical and practical difficulties. For example, it may be argued against evaluation that:

♦ It is liable to be too expensive, and to take funds away from the important programmes that are needed. It is a diversion from the real effort of an Innovation Programme.

♦ It is too difficult to undertake in a serious and systematic manner.

♦ It faces too many obstacles in terms of the reluctance of people to be evaluated, which may mean their being found at fault and criticised.

Another line of argument questions whether evaluation is really necessary, suggesting that it is simply a diversion of efforts. Not only does it contribute little of value, it steals time and resources that could be better used in the programme itself. Thus, we hear suggestions that:

♦ Policymakers and programme designers have acquired a good sense of what works and what does not, and there is little new an evaluation can tell them.

♦ Those evaluated – programme managers, participants in the programme, and so on, are too busy to be diverted into spending time involved in an evaluation.

♦ People who have already been involved in earlier evaluations (or earlier stages of a real-time evaluation) may be suffering from “evaluation fatigue”, meaning that their willingness to answer questions and provide information is eroded.

♦ “Evaluation fatigue” can also affect the intended users of evaluations, who may find themselves bombarded with evaluations if a number are being generated at the same time. With no synthetic overview of the evaluations, they are not likely to be really used as intended.
3.3 How can the arguments against evaluation be countered?

The best way to counter arguments against evaluation is to be able to demonstrate that evaluations can be accomplished and used well. This may be done by reference to successful and useful experience of evaluation in other policy fields. But we are used to hearing that innovation policy is particularly difficult and long-term, which may mean that the relevance of evaluations of policies designed to have more immediate impacts is rather doubtful.

What can be done when there are no evaluations of relevant long-term policies, let alone of Innovation Programmes (or even of research programmes), that can be called on to make the case? In such instances, it may be a matter of drawing on experience in other countries. Pilot studies and evaluation work undertaken by Universities or other parties without government funding may also be called into play.

But there are some generic counterarguments that can be applied when the case against evaluation is being made. This Guide should provide much of the substance to underpin these counterarguments.
3.4 How can the arguments about the difficulty of evaluation be countered?

*Evaluation is liable to be too expensive, and to take funds away from the important programmes that are needed. It is a diversion from the real effort of an Innovation Programme.*

A programme which fails to achieve its objectives, or achieves them in an inefficient way is a waste of public finance. Without any evaluation, the effectiveness and efficiency of a programme cannot be assessed. The costs of a programme evaluation are rarely more than a few per cent of the total programme costs. Evaluations should be able to provide information that will enable more efficient allocation of resources in the future, meaning more funds for the activities that are most effective. Furthermore, well-performed evaluations are important ammunition to use in the effort to gain resources from financial authorities. Evaluations should give programme designers and managers an opportunity to look up from their immediate work to take a more strategic overview of what it is that they are seeking to do - a chance to make sure that short-term concerns are not leading to a diversion of effort that will be problematic for long-term goals.

*Evaluation is too difficult to undertake in a serious and systematic manner.*

This Guide contains much information about how to make evaluations serious and systematic affairs. But evaluation is about improving our knowledge, perfecting it is not possible. Even an incomplete evaluation can yield useful intelligence, though it needs to be explicitly recognised that it is incomplete, so that decisions are not solely founded on a particular subset of performance indicators, for instance. Evaluations can be easily built into programme and policy development and may be developed incrementally from a simple set of tools and methodologies. Much expertise and advice on how to conduct evaluation is available outside of this Guide.

*Evaluation faces too many obstacles in terms of the reluctance of people to be evaluated, when they fear being found at fault and criticised.*

These concerns are often expressed when evaluations are first being introduced into a country or policy area. In general, evaluations are not a matter of looking for individual failures, but attempting to draw systematic lessons about how to perform well. Evaluation is not a “stick” or a means to judge those involved, it is intended to provide benefits to all programme stakeholders, to ensure that benefits are maximised and that no-one is disadvantaged, for example by undue administrative costs. Of course, it may emerge that programmes were poorly designed and that particular activities were not implemented well, and this may reflect poorly on the individuals involved. Knowledge that an evaluation is going to be held in due course then can constitute motivation for better performance! Methodologically, the issue arises that people with something to hide may not be accurate informants for the evaluation team. The implication of this is that wherever possible information should be “triangulated”, that is confirmed by means of several independent sources that are unlikely to be suffering from the same systemic “bias”.
3.5 How can the arguments that policymakers do not need evaluation be countered?

The argument goes:

_Policymakers and programme designers have acquired a good sense of what works and what does not, and there is little new an evaluation can tell them._

It may well be that those who have been associated with many programmes do have a deep understanding of the issues at stake, and a “gut feeling” for whether a particular programme is well-designed and is operating efficiently. Not all Innovation Programmes are so fortunate as to have these wise decision-makers. Even when they do, it is helpful to have external validation of one’s perceptions and suspicions - one of the major benefits of evaluation is that it offers an opportunity for a fresh, ideally independent, view of the effects of implementing programmes. And the evaluation leaves a record of how the programme was undertaken, and to what effect, that can help inform and upgrade the understanding of future decision-makers, who lack this experience. Evaluations can benefit considerably from having the views of well-informed managers about what issues and effects might be particularly worth examining. In practice, we often find that managers are themselves not completely sure about all features of a programme, and actually welcome the insights that a good evaluation can provide. Evaluations also provide a means to test the underlying assumptions (about the innovation system) upon which the programme was originally designed and offer policy makers and programme managers the opportunity to reflect upon the design of past programmes and draw lessons for the design of future programmes. Lastly, programmes may typically last several years and may see a number of programme managers come and go. Evaluations thus provide a means to circumvent this loss of “corporate memory”._
3.6 How can the arguments that evaluation is a misuse of essential resources be countered?

The argument goes:

Those evaluated – programme managers, participants in the programme, and so on, are too busy to be diverted into spending time involved in an evaluation.

Certainly some time is required of participants. In a well-designed evaluation, this is not normally excessive time, and provides an opportunity, as mentioned above, for stakeholders to reflect on their practice, and to think about the broader programme which they are located within. This is often actually welcomed. Indeed, the people interviewed may also want to use the processes as an opportunity to find out about progress and problems in other areas of the programme. Interviews and especially questionnaires are less onerous when well-designed (see below) and allow for easy responses, and do not demand that people simply reiterate material that is in the public domain already. An evaluation also confers a much greater level of programme transparency to all its stakeholders. Besides, participating in or running a sub-optimally performing programme can also be a major waste of resources.

The burden of an evaluation also depends on how evaluations are conducted: ideally, routine data and information should be collected through low-cost and low-key monitoring or “continuous evaluation” procedures which allow the evaluation to focus on key issues of concern to participants. As the outcomes, effects and impacts of programmes would be expected to change and develop over time, different stakeholders should be engaged at different stages of the evaluation.
3.7 How can the arguments about the (lack of) necessity of evaluation be countered?

The argument goes:

Key stakeholders who have already been involved in earlier evaluations are liable to suffer from “evaluation fatigue” – which can also affect the intended users of evaluations, meaning that they are not really used as intended.

Evaluation fatigue can be a real issue when there are numerous evaluations of numerous programmes being undertaken routinely. It can be especially problematic for programme managers who happen to be in charge of a number of programmes following roughly the same timetable. There may be a case for limiting the number of evaluations or simplifying the process (so that, for example, material can be gained on several programmes simultaneously).

Thus, evaluations should be timed to maximise their impacts, perhaps one or two examples from a series of related programmes could be used rather than a full set of evaluations of all the programmes. Different aspects of similar programmes can also be evaluated at different times. Policymakers can also consider portfolio evaluations, which look at the effects of a series of linked programmes rather than examining each one individually. This allows a more holistic view and may identify bottlenecks or synergies between the components of the evaluation system.

This problem (and those mentioned earlier) does not make a case for abandoning evaluation. What it tells us is not to overload people with too many unintegrated evaluations. This last conclusion is one that has resonances across all the criticisms of evaluation discussed above. None of them make a really strong case against evaluation. What they constitute is a set of arguments that cumulatively reinforce the messages:

♦ Evaluations need to be done seriously and systematically: poor evaluations provide little opportunity for learning, fail to demonstrate their benefits and will do little to help build a culture of evaluation

♦ Evaluations are not panaceas for everything that is wrong in innovation policies (or indeed, the national innovation system). They provide an opportunity for learning and as such are important tools in the effort to achieve progress here. But there may be political and other reasons for it being hard to implement satisfactory evaluations, or to see their results properly.

Having considered some reasons why evaluation might be regarded as unnecessary or counterproductive, let us consider the arguments positively for the evaluation of Innovation Programmes.
Arguments for evaluation

3.8 How can we argue for evaluation?

Innovation is a multifaceted phenomenon, which policymakers are approaching from many angles. There are thus many different roles that evaluation can play in relation to Innovation Programmes. There are several different views about these roles, too. None is necessarily the correct view – all depend on the context in which evaluation is being used and what lessons policy makers wish to gain from its use.

When the case for innovation programmes is seen simply as a matter of market failure, for instance, then the main task for evaluation is primarily to see how far this market failure has been addressed – are information or training gaps being overcome, for example? When the problem is seen more in terms of innovation systems and their performance (the notion of systems failure, for example), then evaluation has to address how far systemic change has been achieved, i.e. through the removal of bottlenecks, or through the improvement of framework conditions, etc. This may involve questions of the change in behaviour of economic agents over the long-term.

The arguments for evaluation are, then, equally multifaceted. It is possible to identify a number of major reasons that are given for undertaking Innovation Programme evaluations (and evaluations more generally) 3, and also a number of more minor or incidental ones. We shall consider each of the following in turn on the coming pages:

♦ the rationale based on assessing value-for-money
♦ the rationale based on improving the design of future programmes
♦ the rationale based on informing the priority setting process
♦ the rationale based on enhancing policy design
♦ other rationales and benefits – for instance, use of evaluations to further disseminate programme results.

---

3 EPEC (2005) presents a similar range of rationales for evaluation, assessing how far each type of impact was attained across a set of evaluations.
3.9 What are the arguments around ‘Assessing value-for-money’?

Simply, the assessment of value for money is a matter of asking the question of how well public money is being spent. This is a fundamental question – government expenditure is not a bottomless pit. Money has to be extracted from taxes and other revenues, and applied over a range of policy goals. It is only ethical to ensure that it is being used well: otherwise, what right does the state have to this money in the first place? Ministries of finance and public expenditure watchdogs/auditors are the institutions that need to be convinced that money is spent well, and the mass media and political opposition may be interested in whether this is the case, too. Evidence of value for money is important if a case is to be built for subsequent programmes of this kind. And, when carried out as part of a real-time evaluation it may help decisions about whether to change management practices in the programme and even whether to modify the allocation of resources across activities.

The sorts of questions that follow from this ‘value for money’ rationale are:

- How far have the proposed activities actually been carried out?
- If they have not all been carried out, are the reasons for this adequate ones that could not have been foreseen in the programme design stage?
- Were the activities carried out efficiently, i.e. with the minimum of administrative overheads and management costs?
- Were the procedures for reallocating resources in the event of discovery of problems with one or other part of the programme adequate? Were they carried out efficiently?
- How far did the activities achieve their intended benefits and what is the ratio of benefits to costs, as far as possible expressed in financial terms?

Some of these are essentially auditing questions that should be valuable on the basis of information collected routinely during programme management. In this sense, it is useful for programme managers to collate data on whether milestones are being achieved, on the uptake of funds, and whether reports and other outputs have been received on time. It may be that non-routine information has also been collected and processed – for instance, that a particular expenditure was not incurred because the firm involved closed down or moved operations, because initial piloting indicated that the intended activity was not at all effective, and so on. Often such information is not written down, or if it is it will need to be checked and amplified. Likewise, the procedures for reallocating resources in the event of such contingencies will need to be explicated, and the way they were implemented is probably a topic for enquiry.
3.10 What are the implications of Evaluation goings beyond Auditing and Monitoring for arguments about ‘Assessing value-for-money’?

Auditing procedures are ways of checking whether programmes have been conducted as intended, and that money has been spent on planned allocations with due diligence. Auditing is essentially a verification process. However, as is the way with many terms and practices, the scope of “auditing” has grown over the years. On the one hand, we may have “performance audits”, which see how far targets have been achieved in terms of given performance indicators. On the other hand, auditors may have to take into account the level of knowledge that was available at the time a programme was planned and implemented, to establish whether decisions were taken on the basis of (apparently) well-founded intelligence. These activities bring auditing closer to evaluation, but it still tends to remain a narrower and more descriptive process. In essence, auditing generally focuses on inputs and their allocation and may extend to an examination of the process of input utilisation. What it fails to do is to examine outputs and outcomes, and the processes and reasons that generate them. Evaluations typically go beyond audits in being more analytical, aiming to draw lessons about the planning and conduct of Innovation Programmes more generally. This is not to say that evaluations ignore inputs, rather they seek to examine the links between programme rationales, inputs, activities and outcomes.

A similar point can be made in respect of monitoring. Monitoring is the process of examining programme activities during the course of the programme itself, and thus may appear to resemble “real-time” evaluation. But monitoring is more or a routine management activity. It is a matter of checking to see that activities are proceeding, and funds being allocated and used, as planned. It does not, of itself, address the wider evaluation issues concerning the objectives and broad effects of the programme. However, this is not to denigrate the role of monitoring which may be used to accumulate, on a routine (and possibly low-cost) basis, information essential to the conduct of the eventual evaluation. It is frequently the case that evaluators are required to carry out resource-intensive data gathering exercises to acquire information which might have been collected routinely during the course of the programme. As we shall see, early provision for the identification and collection of this information should form part of the programme design process.

The ‘value for money’ rationale has a strong focus on expenditures: how they are managed, how they relate to outcomes. The processes examined are those that relate to expenditures, and the main issues at stake are whether public funds are being used efficiently – and if not, what are the shortcomings in the programme design or management. There is little reason to reflect on whether the programme objectives were really the most appropriate ones, or how the activities fit into the wider innovation system.

Evaluation in this case is in large part an auditing process, drawing on routine management information, and needing to go into detail only in understanding deviation from plans and assessing benefits from the programme. With a
pressure to express benefits in financial terms, the sorts of indicators of performance that are chosen will tend to be those that can most easily be represented in terms of money: for example, turnover associated with new products, economic value of new jobs created.

Questions about the achievements and cost-benefit ratios associated with a programme are, typically, more demanding. In research evaluation, those in receipt of funds are commonly required to provide end-of-project reports which contain evidence of publications stemming from the work. The outcomes and effects and, hence, the impacts of wider Innovation Programmes are not usually so straightforward to capture.4

4 Not that a list of publications is always so straightforward. Even if the compilers of the lists use a standard format for presenting them, there are often such problems as: the difficulty in attributing specific papers to a specific project; the likelihood that some planned publications will be abandoned while others appear; and the long delay in preparing publications.
3.11 What are the arguments about ‘Improving programme management’?

Programmes do not just manage themselves. It is well-known that the calibre of the programme managers can play a vital role in ensuring the success of a programme – how alert they are to signals about its progress, how energetic they are in exploring solutions to problems, how effective they are in communicating awareness and enthusiasm to participants in the programme. Obviously, selecting programme managers who have high capabilities, as well as being committed to the objectives of the programme, is of great importance.

Management structures can also vary in terms of their effectiveness. The best managers may have a hard time making progress in the absence of such structures, while the poorer managers will be able to make a better job – and even improve their capabilities – where such structures are in place. This Guide discusses the management of evaluation processes, and some of the points here apply to programme management more generally; the wider management literature (especially – but not only – public sector management material) also provides a great deal of useful orientation to the tools and techniques of good management. Evaluations can contribute in terms of assessing the management practices involved in an Innovation Programme. For example, an evaluation can examine how various procedures were implemented, drawing lessons about how they should be designed in order to achieve best results. For instance:

♦ How were the programme partners selected and mobilised?
♦ How can programme objectives be adequately communicated to the stakeholders involved, especially the programme’s participants (and other targets)?
♦ Were the activities properly specified and monitored? Were there appropriate reporting intervals and guidelines for the information required, and were the management procedures for processing this information and for the detection of problems adequate? How can the progress of activities be efficiently monitored and matched against plans and targets?
♦ Was there adequate awareness within the programme concerning its rationale and objectives, the ways in which activities were meant to relate, etc?
♦ What procedures were in place for the wider dissemination of results of the programme?
♦ How did the programme management respond to contingencies that may have arisen, for example, how were resources reallocated in the event of differential progress toward objectives?
♦ How should the evaluation process itself be managed? What procedures were in place to prepare and facilitate the evaluation process, what procedures should be in place to ensure adequate follow-up of the conclusions and recommendations from the evaluation process?
3.12 What is the role of Evaluation in ‘Improving programme management’?

General management principles, even if learned from experience in other types of public programmes, will always need some degree of refinement to fit the specificities of particular policy areas. Innovation Programmes have features that make them different from most other public programmes. These involve the types of:

♦ behavioural changes that are being sought
♦ impact that these changes might generate (over lengthy time periods and broad areas of economic life)
♦ economic actors whose behaviour is intended to be affected, and
♦ socioeconomic networks in which these actors are located.

The management procedures and processes involved in Innovation Programmes are thus by no means standard ones. Evaluation can play an important role in helping to identify and explicate the main management structures and capabilities that are required in relation to particular types of programme.

Each evaluation undertaken by an institution will make a further contribution to that institution’s knowledge-base, in terms of its understanding of the innovation system, the actors and processes within it and the institutions capability to plan, implement and assess the policy interventions it makes. The initial, pioneering evaluations, if they address programme management processes, will contribute to establishing a basic understanding about the principles of effective management of Innovation Programmes. As these principles come to be well-understood, these lessons will tend to become repetitive, though this may be precisely what is required for the lessons to be learned and the understanding of the principles to be embedded into the operating agencies in question. There may be some tendency for later evaluations to become more of an assessment of whether best practice is being followed in management terms, with less contribution of fresh insights into management principles.

However, management disciplines are continually evolving, which means that there is always scope for refreshing the principles and processes that are being applied. Models and methods may be imported from the private sector (e.g. “lean management”) or other public bodies (“new public management” approaches, for example), new technical tools may enable approaches that were not viable earlier (in particular, new Information Technology can be used widely for project management, co-ordination of activities, dissemination of results, etc.). As evaluation is a powerful learning tool, we should therefore expect practices of Innovation Programme management to be continually evolving, thus the evaluation of programme management can be expected to keep moving too.
Such evaluation is, again, more than just a matter of ticking boxes to determine whether best practice was being followed in various elements of programme management. It should be oriented to learning lessons from programme management, which can be used to inform the design of management structures and processes for future programmes. The goal should be to attain continual improvement in Innovation Programme management, which will increase the positive outcomes of the programmes and help build the cases that public funds are being well spent on vital objectives.
3.13 What are the arguments about ‘Improving design of programmes’?

Evaluation in terms of ‘value for money’ can tell us whether a programme proved to be worth the funds expended on it. Evaluation of management procedures tells us a great deal about how efficiently funds were being used, and whether more value-added could be achieved by improving management processes. Good evaluation of this sort should be also able to identify reasons behind any shortcomings, and thus help inform future programme design.

But the design of future programmes can and should be a key part of the rationale for evaluation. Evaluation thus becomes a tool to enhance learning about which interventions have what consequences, and why, and how to effectively design future interventions that will really contribute to the major policy goals being pursued.

Therefore, evaluation is far more than something that is just ‘bolted on’ at the end of a programme, to satisfy its sponsors that the funds have been well spent. Instead, evaluation should be seen as an integral part of an individual programme. Moreover, it is an integral part of the wider innovation policy process, which can and should play a role in the design of programmes more generally.

Programme design is closely related to goal setting. The design of any programme will be just one of a huge range of alternative programme designs available to policy makers, some of which may have been explicitly investigated and considered, some of which may not have been articulated.

Programmes may differ in terms of their ultimate goals. This is particularly true of innovation programmes. Examples of different ultimate goals could be, for instance, to: improve innovativeness across the economy; leverage regional development though more innovative clusters in disadvantaged regions; encourage industry to explore more energy-efficient lines of technology adoption; and so on. And, of course, there are higher-order policy choices between the allocation of resources to goals such as national security, immediate public health, environmental sustainability, economic stability, and other critical features of a well-functioning society. In this discussion we focus on the major goals behind Innovation Programmes.

Within a shared ultimate goal, there is a choice between different types of activity that could contribute to this goal. Examples of different types of activity could be, for instance, efforts to: raise awareness; support technology adoption (e.g. through funding consultancy support for firms or clusters); encourage better linkages between researchers, technology developers, and technology users; increase quality or quantity of training for technology use; and so on. Within any activities of this sort there may also be sub-activities: how far to support each of a number of relevant technologies, types of training programme, and so on.
Different programme designs reflect choices that have been made between these alternatives. Making such choices is a matter of setting priorities, which may be accomplished explicitly or more or less implicitly.

The role of evaluation in this context is to assess which goals are achievable, and by what means they may most effectively be accomplished. In short, evaluations are used to assess the consequences of policy action. But the way in which evaluation is to be used needs to be considered from the outset of the policy design process. This will facilitate the measurement of these goals and assessment of how far the programme has contributed to achieving them.
3.14 What is the role of Setting Objectives in ‘Improving design of programmes’?

Choices in terms of ultimate goals are typically made at high political levels, and it is hard for a single evaluation to contribute much to them. However, some large-scale evaluations of large-scale programmes have proved very influential. Furthermore, a series of evaluations may be used as a knowledge base from which advice may be derived as to which policy goals may be achieved fairly readily, and which are more intractable, at least in terms of the instruments that have so far been brought to bear on them. This ideal is usually only realised imperfectly, if at all, due to the uneven development of evaluation culture across different policy areas, and thus the uneven quality and approach of the evaluations that are available. Knowledge-based policymaking still has a long way to go! Nevertheless, the availability of a “library” of innovation policy evaluations can serve to inform policy makers and programme managers alike about the broad areas of intervention that are more likely to succeed - and, equally important, about the main problems that are likely to be encountered with their use.

Choices in terms of the objectives of a programme, and especially choices about the activities and sub-activities that are instituted as ways of achieving these objectives, is usually less a matter of high policy and politics. Here, evaluation work can contribute to spelling out the relative efficiency and effectiveness of different activities and sub-activities in contributing to programme objectives. The Programme Overview should have spelled out not just what these objectives are, and how specific activities are meant to contribute to them, but also have gone some way to explicating the choices that were made about specific objectives and activities. The evaluation can review these choices in the light of experience, and provide intelligence for the choices required in future programmes.

The question for each individual evaluation is, in this light, a matter of what the experience of the particular programme (or suite of programmes) being evaluated tells us about how to design future programmes. The design of the programme encompasses two dimensions. First is the dimension of programme content – where evaluations can contribute to learning about setting and specifying objectives, relating these to activities, and so on. Second is the dimension of programme structure – where evaluations can contribute to learning about effective monitoring of such a programme, about the management capabilities and practices required to see it through successfully. This second dimension is discussed above, under the rubric of “improving programme management”. Evaluation here should help with the design of programme management processes once the strategic decisions associated with the first dimension have been taken. The first dimension involves selection and design of programmes that are appropriate activities and uses of resources in light of policy goals. Evaluations can also contribute intelligence that should help with these selection and design processes. In so doing, they will be upgrading the policy institutions’ understanding of the innovation systems that their policies are seeking to affect.
3.15 What is the role of Evaluation in ‘Improving design of programmes’?

As noted above, the easiest contributions for evaluation to make are those that concern programme objectives, activities, and sub-activities. The evaluation can tackle such topics by asking, for example:

♦ What lessons can be learned about the rationale and objectives of programmes? Was the rationale correctly conceived? How relevant do the programme objectives now appear to be in terms of contributing to the overall innovation policy goals? Were there superfluous, unrealistic or missing objectives?

♦ Has the evaluation suggested that certain objectives are particularly easy or difficult to achieve? Are some objectives too ambitious, now that we understand better what the effects of interventions are? Do particular objectives require different or complementary approaches? Can the achievement of objectives actually be assessed?

♦ How far do particular activities and sub-activities contribute to the objectives with which they are associated? Does it emerge that there are activities that are less effective than others; have other potential activities been identified which were not in the original programme?

♦ What does this imply about future programmes? How should the balance of objectives, and of activities and sub activities, be modified? Are there promising directions for further work that should be explored?

Even though it is unlikely that successive programmes will have exactly the same objectives, the evaluation of one programme can be an opportunity to consider where the priorities should lie in the next programme. Successive programmes may vary in more or less incremental ways, for instance placing more emphasis on reaching particular types of firm, of promoting the adoption of particular types of technology or technique, etc. The evaluation may suggest that programme design needs a new approach, to draw on inputs from a wider range of expertise or from other studies and reviews, or to be constituted so that there can be better learning from the results of earlier and ongoing evaluations.

What does this imply about evaluation?

This rationale means that the evaluation should examine programme objectives, and how efforts are made to realise them, in the light of the broader innovation policies to which they are intended to contribute. While it may be hard to make a simple quantitative calculation here, the underlying issue is whether the objectives (and the activities that follow from them) actually do constitute the sort of operationalisation of the broad policy goals that were originally envisaged. In the light of experience, have any mismatches been found between the translation of policy goals into programme objectives?
This also implies that the participants in the programme should be asked to reflect on the programme more generally, in addition to those activities with which they have been involved, and to think about how programme objectives might be modified and/or better addressed in any successor programmes. Thus, for example, during the course of the evaluation, the evaluators may ask programme participants what they consider to be potentially fruitful directions for future programmes to take, and why. These views can inform advice on future priorities for activities within related programmes.
3.16 How can we argue about ‘Enhancing policy design’?

The most ambitious role for evaluation is in contributing to the mix of policies that are brought to bear on the enhancement of innovation. This is a function that will usually require more than one evaluation study, though even a single study can provide valuable intelligence.

Here, the purpose of evaluations is to learn deeper lessons about how and why programmes have achieved their particular pattern of results – understanding how the programme activities have operated in and impacted on particular elements of the national innovation system. Evaluations thus use Innovation Programmes as a form of ‘action research’ into the innovation system, using the programmes to determine features of this system. Evaluations, via innovation programmes thus pose questions such as where are the points of weakness, in which areas can leverage be obtained, why it is that some interventions seem to be far more effective than others, and so on?

Hence evaluations are embedded into a policy learning process, wherein decision-makers assess the mix of policies they are putting into place. As a result of accumulating deeper understanding of how the innovation system is functioning policy makers are able to design new policies and change the emphasis given to specific types of policy. One of the main benefits of such an approach to evaluation is that it can help identify where policies are failing to complement each other, and even where there are serious mismatches. For example, education or competition policies may operate in ways that tend to militate against policies for, say, small firm innovation.

What does this imply about evaluation?

This rationale requires good evaluations, of course, but more significantly it implies that a systematic culture of evaluation is being built. Multiple evaluations are being conducted, and the results of evaluations are being widely examined and used. Users are able to evaluate these results, and integrate the findings of different evaluations. There may be a shift to complementing evaluations of single programmes (which will still be necessary, especially where new types of programme are involved), to more of a “portfolio approach” (Polt, 1998) that sets out to examine and draw conclusions from a range of programmes. The implication is also that a range of methods should be applied because this is necessary to establish the range of knowledge about the innovation system. In particular, it may be necessary to investigate “softer” elements of this system, such as the networks that exist between its constituent actors.
3.17 What other arguments are there for evaluation?

There can be several additional reasons for pursuing an evaluation, for instituting evaluation practices and building an evaluation culture more generally.

An individual evaluation may play several roles in addition to those identified in the rationales discussed earlier. It may:

♦ Document the programme and its achievements in ways that are useful for wider dissemination of the results. The evaluation report may be published by the programme itself or by the programme sponsors as a record of the activities and outcomes. The evaluation may even engage in “dissemination” as a by-product of evaluation activities - for example, asking stakeholders about the effects of activities may require informing them about the full range of activities that formed part of the programme.

♦ Help in promotion of the underlying objectives of the programme. The evaluation should provide an opportunity to re-examine the programme rationale and objectives, restating these in the light of experience and concrete evidence. This can be a timely input to debate about future programme objectives. Evaluations are not the same as “success story” approaches (see below), which are explicitly designed with promotion of the programme and its objectives, and dissemination of some of the programme results to a wider audiences – although they may assist in the identification of “good practice” examples and cases. The account of the programme from the third-party perspective that an evaluation provides can be a useful product in its own right for stimulating discussion of the interventions that were undertaken.

♦ Promote evaluation itself. The study can help build an evaluation culture, by exposing various parties to the practices and results of evaluation in this policy area.

♦ Benefit later evaluations, by providing a good example of what can be done, by developing templates for interviews and questionnaires, by establishing a record of achievements and conditions that can be used as a baseline in future work.

More generally, efforts to build an institutionalised culture of evaluation have been discussed in terms of the contributions that this can make to improving innovation policy. It may also:

♦ Help make the case for evidence-based policymaking in other spheres of government, and lead to better appreciation of the skills and procedures that are necessary to support this

♦ Help to stimulate evaluation and related activities among stakeholders and actors in the innovation system more generally, hopefully helping them to
move to more strategic views and approaches in their efforts, and putting innovation itself more on to the agenda.\(^5\)

**What does this imply about evaluation?**

These various features are more in the nature of incidental and additional benefits than a full rationale for conducting an evaluation. *Most of these benefits can only be realised if the evaluation is conducted transparently, with well-documented methods and results; and the approach adopted and its conclusions need to be made publicly available.* The work will need to be conducted and presented in such a way that it – and its relevance - can be understood by the numerous stakeholders and interested parties involved in the system.

\(^5\) Eiffinger (1997) describes how the Dutch research evaluation system was deliberately set up to encourages stakeholders such as Universities to undertake their own evaluations.
3.18 What, then is the overall case for evaluation?

The rationales discussed above cover a broad range of goals for the evaluation exercise. The various goals are quite complementary – pursuing one goal need not rule out pursuing others. But in any particular evaluation, there is bound to be some emphasis on one or other approach. Indeed, it is likely that an institution’s evaluation regime will tend to be oriented towards a specific rationale, rather than allowing each evaluation to take a completely different form. We discuss later some of the main types of evaluation culture that can be encountered.

Whatever the specific circumstances, several general points arise from the preceding discussion. First, there are strong cases for evaluation to be a regular and routine element of Innovation Programmes. The more that this is an accepted part of programme design, the more that design can facilitate evaluation activities by already specifying, for example, data to be generated in the course of the programme activities, the objectives and targets against which evaluation can be conducted, etc.

Second, there are potential benefits from evaluation that go well beyond the important function of ensuring public funds are being well spent. Evaluation can contribute to a more informed process of programme design, and it can be a vital element in a process of policy learning. Ultimately, evaluations are a way for policymakers to gain better understanding of the system – in this case the innovation system – that they are trying to influence, and the ways in which interventions of different kinds, at different times, can effect this system.

None of these arguments seek to elevate evaluation to a primary role in the policy process. It is a vital element of this process, but one element among others. It is not an end in itself. There can be too many evaluations to be thoroughly assimilated. There can be emphasis on evaluation at the expense of the activities that are actually being evaluated. There can be over elaborate efforts to understand the “complete” range of outcomes of a programme at the expense of providing timely (or even the most relevant) information for policymaking. Especially if carried out in a rigid fashion, with strict adherence to a narrow set of targets and indicators, the evaluation process can come to limit the conduct of programmes. The arguments against evaluation discussed earlier raise serious issues that need to be addressed adequately in individual evaluation studies and in building institutional evaluation systems and an overall culture of evaluation. There are pitfalls from too much and too rigid evaluation, and there are inherent limitations of the evaluation approach (i.e. the risks of overambitious evaluation). Recognising these does not diminish the case for evaluation. It simply tells us more about how we should frame and conduct evaluations.
Chapter 4 EVALUATION APPROACHES

4.1 How does Evaluation differ from Monitoring?

Monitoring is a process of providing management information through the life of a programme. It may be assisted by project management and similar tools. It comprises systems for obtaining signals as to whether activities are taking place according to schedule, if milestones are being met, whether funds are being used for the intended purposes, and the like. The focus tends to be on inputs and activities, and less on outputs and outcomes, though immediate outputs like project reports and other deliverables are encompassed. Monitoring is usually carried out by the programme management team, who will react to any information that things are not going to plan. Evaluations are usually undertaken by independent parties, who seek to draw wider lessons from the programme.

However, monitoring can form an important element within an evaluation and within the overall programme/policy design and implementation cycle. As has been noted, evaluators, particularly those from outside the programme, often have to familiarise themselves with the operational characteristics of the programme. Moreover, they frequently wish to obtain what can be viewed as mundane information on programme inputs (such as resources allocated per participant) and outputs such as publications, reports, patents, etc. Such informational needs should be anticipated in the programme design phase and adequate mechanisms put in place to ensure their routine capture by the programme administration. If such information is then available at the start of an evaluation, valuable time and resources may be spared which otherwise would have been spent in capturing this data. In addition, there is a danger that if programme managers do not capture this information or if participants are not aware of the need to compile this information, it will be lost, either partially or entirely. At best, participants will need to waste their resources in searching for such information in response to evaluators’ queries. The resources thus saved may then be channelled into more illuminating lines of enquiry.

For this reason, monitoring procedures should also be considered when designing the evaluation elements of a programme. Of course, there is also a trade off to be made between the type and amount of information collected during the monitoring process and the effort required. Monitoring should have a low administrative burden and the potential value of the selected information (to the evaluation process) should be carefully considered.
4.2 How does Evaluation differ from Auditing?

**Auditing** is a process of checking that a project has been managed as planned, especially that resources have been allocated and spent properly, that activities have been undertaken as planned, that outputs have been forthcoming. It usually performs a check of these things after programme completion and is carried out by parties other than programme managers. There may be annual or mid-term auditing processes, however, and sometimes these will be required in order for the remainder of the funds to be provided. Auditing is less focused on explanation and analysis of the system of innovation (in this policy area) than are evaluations. Auditing tends to be more focused on inputs than monitoring.
4.3 How does Evaluation differ from Analysis of Success Stories?

**Success Stories** can be more or less closely related to evaluation studies. The focus here is on identifying cases of particularly successful activities, projects, or participants in a programme, and explicating the course of events that led to this success. Where the success story is simply a demonstration of the virtues of a programme, or of the significance of its objectives, it is liable to be more descriptive. When it includes more analysis of how it was that the case was so successful, what the preconditions for this were, and so on, it moves closer to evaluation. But unlike evaluation it does not attempt a systematic overview of the whole programme, nor will it explain what went wrong, and why, in unsuccessful cases – evaluation also helps us to learn from mistakes or failures. The success story approach resembles the approach of many management studies in the private sector, which emphasise the identification of best practice and learning from excellence. Evaluation has more in common with conventional social science approaches, setting out to get the wider picture within which individual cases can be understood. The approaches can be complementary – evaluations can suggest cases to elaborate in success stories (and some element of success storytelling can be built into an evaluation); the success story approach can suggest key variables for the evaluation to examine. There is not inherent conflict between the approaches, since they serve slightly different purposes. However, they should not be mistaken for each other.
4.4 What are the Main Types of Innovation Programme Evaluation?

Evaluations differ in all sorts of ways – in the methods used, in their scale, their scope, and in the extent to which results are disseminated and used. But there are three fundamental types of evaluation, which differ in terms of their timing, and which it is particularly important to differentiate:

- **Ex-ante evaluation**, conducted before the implementation of a programme, and focusing on what its objectives are and how (and how far) they should be achieved.

- **Intermediate evaluation**, reviewing the progress of the programme, or its achievements at some point into the programme – usually reserved for long-lasting programmes.

- **Real-time evaluation**, following the programme in detail through its operation.

- **Ex-post evaluation**, examining the results of the programme after it has been completed (and possibly more than once).
Figure 4.1 Types of evaluation

<table>
<thead>
<tr>
<th>Period before programme</th>
<th>Programme duration</th>
<th>Period after programme...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formulation</td>
<td>Stage 1</td>
<td>Stage 2</td>
</tr>
<tr>
<td>Stage 2</td>
<td>Stage 3</td>
<td>Conclusion</td>
</tr>
<tr>
<td>Conclusion</td>
<td>Follow-up</td>
<td></td>
</tr>
</tbody>
</table>

Types of evaluation:
- Ex Ante Evaluation
- Interim Evaluations
- Ex Post Evaluation

Real Time Evaluation
4.5 What is *Ex Ante* Evaluation?

There is some doubt about whether *ex ante* evaluations should really be referred to as evaluations, rather than being described by some term such as “appraisal”. This would no doubt reduce the scope for confusion. But since it is established practice to refer to these activities as evaluations, we shall continue to do so here – while noting that the real-time and *ex post* evaluations described below are able to provide more evidence based on actual experience of the programme itself.

As the phrase suggests, *ex ante* evaluations are conducted **before** the programme is implemented. Indeed, they typically inform the Programme Overview and any ROAME statement that has been prepared. In terms of the objectives of the programme, the *ex ante* evaluation brings together evidence and arguments concerning the likely consequences of the programme activities. Since this will involve a process of matching activities with objectives, this evaluation may challenge the clarity with which the programme objectives have been set out, as well as the plausibility of the objectives themselves. This may even extend to examining whether the objectives are realistic and whether they represent adequate contributions to the broad policy goals that underpin the programme. EPEC (2005) notes, however, that *ex ante* evaluations tended to have most influence on details of programme design, and for the implications to be more incremental ones.

Undertaking such evaluations is a key part of the embedding of evaluation processes into programme design. It is a major element of an evaluation culture in another way, too. *Ex ante* evaluation provides a valuable input to later – real-time and *ex post* – evaluations. It should have created a knowledge base (a “virtual benchmark”) against which these evaluations can be undertaken. It is often said that good *ex ante* evaluations are major contributions too the achievement of good *ex post* evaluations.

---

6 It would reduce confusion about what “evaluation” means, but “appraisal” carries its own set of meanings. *Ex ante* evaluation can be confused with “project appraisal”, for instance, which is usually taken to refer specifically to the selection process across a range of different projects that are being considered.
4.6 How can *Ex Ante* Evaluation Work?

In the absence of any actual results from the programme activities (which, obviously, have yet to be undertaken), *ex ante* evaluations must work through the programme rationale, objectives and activities, and explicate the cases first, for public intervention in terms of the market and/or system failures being confronted, and second, for the specific objectives and activities in terms of the anticipated benefits that should be achieved by the programme. This may involve two components:

♦ Modelling: This involves a formal approach to estimating the likely magnitude of impacts and quite possibly quantifying these in terms of economic benefits. For instance, models may be used in constructing accounts of the numbers of firms that might be reached by a programme and the amount or type of change in their behaviour that could be expected. They can be used to consider the upstream and downstream implications of such behavioural changes on the suppliers and customers of the targeted firms (these are often more significant than the impacts on the firms themselves, in terms of job creation, for instance.) Models can be very simple or highly sophisticated. In the case of programmes operated by an agency or body that undertakes a number of programmes at any one time, it may well be that there are established models (e.g. based on cost-benefit, input-output or econometric approaches) that can be used.

♦ Scenario analysis: This can have two functions. The first is to contrast the likely impacts of the programme with the “counterfactual” situation where the programme does not happen. This constitutes part of the justification for the programme and may provide some basis for the estimation of the programmes “additionality”. The second is to consider alternative environments which may arise during the course of the programme’s operation, and which might influence its results (or even affect the need for intervention). This contributes to contingency planning and to developing procedures to deal with unexpected issues. It may also lead to some reappraisal of cost-benefit expectations, based on the recognition that the “most likely” scenario may not be all that likely, and that elements of other scenarios are also possible features of the programme’s future.

We discuss methods for *ex ante* evaluation in more detail in chapter 6, below. *Ex ante* evaluation involves a degree of forecasting, and is usually bound to get some things wrong. This is part of the learning process, of course. We can come to understand why and how far expectations were unrealistic ones, necessary procedures were not thought of in advance, and so on. This can help inform the design of successor programmes. However, it does serve to question the rationale and develop a deeper understanding of the possible implications of a programme than might be obtained through the comparatively simpler process of formulating a direct policy response to a specific problem.
4.7 What are Interim, Periodic and Real-Time Evaluations?

These are two terms that are sometimes confused. By “interim” or “intermediate” evaluations we refer to evaluations that are conducted some way into a programme, at a point where some early results should be apparent. This sort of evaluation is most often conducted when a programme has been established to run over a period of several years – four or five or even longer – and when there are a succession of programme activities or distinct phases. The evaluation can inform the implementation of later stages of the programme, the selection of activities and sub-activities, and so on. It can play an important political role in providing legitimacy for ongoing long-term funding. “Periodic” evaluations are typically planned as a series of evaluations to be conducted at specified points in the programme life time. They may be conducted on a regular basis (for example, every five years) or timed to coincide with events in the wider policy cycle (for example, large-scale funding reviews, budgetary cycles, etc.). They share many of the features of ex post evaluations in terms of the methodologies employed and being essentially backward looking and focused on historical evidence, although the programme is still on-going.

“Real-time” evaluation refers to something rather different and rarer. This is an evaluation process that is built into and alongside the rolling out of the programme. Here, evaluators are typically interacting with programme participants as the activities are being undertaken, they are attending programme events, and are able to observe what is happening (rather than needing to reconstruct it subsequently from records and recollections).

The process of real-time evaluation is an extensive one – typically this is a process that goes on through the life of the programme. It is accordingly labour-intensive and potentially rather costly. Thus it tends to be employed only when there are very large programmes underway – and the associated risks of failure are greater. The evaluators may be seen as agents of management or of outside political forces, and their presence be resented by programme participants. On the other hand, a danger with such evaluations is that the evaluators may themselves come to be too close to the programme, to be “captured” and thus to find it hard to make an independent appraisal. However, in large programmes there are usually enough competing points of view that this is less of a problem.

The major advantage of real-time evaluation (shared to some extent with intermediate evaluations) is that it can provide early warnings and timely information that can alert and inform programme managers of the need for action. Going beyond routine monitoring this can, for example, highlight where participants are experiencing difficulties, indicate where future underperformance might be expected, illuminate unexpected consequences of activities, identify where there are shortcomings in communication or challenges to morale, and so on.

All three of the above approaches are quite different from the other real-time process of monitoring that should be routinely accomplished during the
course of a programme. Monitoring is closer to a management process of checking that funds are being spent, activities accomplished according to time and that information on certain immediate outcomes or outputs is being captured. While its major concern is efficiency, it nevertheless forms an important element of evaluation, particularly *ex post* evaluation. Evaluations address, in more depth, issues of impacts and outcomes, and seek to gain more insight into what is taking place. Their concerns go beyond efficiency to include effectiveness, efficacy, and explanation, but they still may benefit from well-formulated monitoring practices.

We discuss methods for real-time and interim evaluations in chapter 6, below.
4.8 What is Ex Post Evaluation?

Ex Post evaluations are undertaken after a programme is completed, or at least near enough to the end of the programme that most activities can be examined and the more immediate outcomes assessed. It may be that the evaluation is started before the programme is formally completed, but goes on for some time after other activities are wound up. Sometimes the evaluation is, along with the reporting of the programme, the last activity in the programme schedule.

Ex post evaluation probably constitutes the bulk of evaluation work that is undertaken nowadays. It allows for the intentions of a programme to be confronted with the realities that the programme activities have encountered and for questions about the effectiveness, efficiency and efficacy of the programme to be addressed.

We discuss methods for ex post evaluation in chapter 6, below. The methodological advantages of ex post evaluation are apparent: only after a programme is completed (or at least in its late stages) is it possible to make a reasonably thorough accounting of its impacts and outcomes. However, there are some disadvantages of the approach:

♦ First, the intelligence that is yielded is obtained too late to be used to inform the management or administration of programme itself, to be used to make changes that could render the programme more effective.

♦ Second, there are similar problems in terms of its possible contribution to successor programmes. The time taken to complete the evaluation may be problematic in terms of the policy cycle where one programme might be expected to follow another immediately. It can also be problematic in terms of the need to maintain activities, energise networks, and the like, when one programme is accepted to be only one of a number of necessarily successive initiatives. Evaluation results may either arrive too slowly to fully inform the design of the next programme, or there may be a gap established before this programme design is undertaken, making for a hiatus in the activities.

For these reasons, some real-time evaluation is often welcome as a means of obtaining early warning on major lessons and conclusions from programme evaluation. It is thus necessary to ensure that the design of the real-time and ex post evaluations is well-integrated, so that the data collection and other work of the former can be used by the latter – and does not get in its way (e.g. by creating “evaluation fatigue” among programme participants). A similar role may be played by efficient monitoring practices wherein data collection can, to some extent, be made a routine procedure and less of a burden on programme participants.

---

7 Sometimes an evaluation will conclude that there is an urgent need for further work to rapidly build on the achievements of a programme. For example, it may be determined that new networks have been established that are extremely promising, but that are liable to fall apart and need rebuilding from scratch, if there is not rapid input of further resources to motivate continuing joint activity.
4.9 What does it mean for Evaluation to be an integrated part of policy making?

Innovation has become recognised as a phenomenon that goes beyond the adoption of individual innovations by individual firms, and goes beyond the innovativeness of individual firms to encompass such things as the systems and networks within which these firms are embedded. Hence the scope of programme evaluation has grown. Innovation programmes are attempts to affect these systems, and to allow for better sharing and exchange of knowledge. There is now more rapid awareness of innovative possibilities and how these relate to user demands, and so on. As the pervasiveness of innovation and the scope of innovation systems have broadened, the opportunities for policy intervention have increased and a range of innovation programmes are possible. Evaluation of these various interventions thus becomes a way of finding out about the innovation system, as well as about the effect that the programme has on the system. Therefore, evaluations should provide cumulative knowledge, rather than just one-off insights into specific programmes. This knowledge should be available not just to policymakers – though it certainly should inform policy – but more generally within the innovation system.

One implication of this systems approach is the need for multiple approaches to evaluation. Each evaluation is liable to have different features, reflecting the specificities of the programmes being evaluated. Additionally, any one evaluation is liable to use more than one method and source of data. The range of impacts being sought in most innovation programmes is usually sufficiently diverse as to require several methodological tools; and often a single objective will require “triangulation” across several indicators, since indicators by their nature provide only partial intelligence on the topics of interest,

Thematic evaluations are also relevant here. These are evaluations where many programmes (or all those conducted during a period of time) are evaluated at the same time. There may be a focus on a specific theme – for example how the programmes helped firms to find international partners to enhance innovation – or on a broad spectrum of impacts of the programmes. Thematic evaluations are growing in number in advanced evaluation cultures.
4.10 How does Innovation Programme Evaluation differ from Research Programme Evaluation?

As noted before, research programmes aim at the promotion of research (and development) activities, generally in a fairly straightforward manner through the provision of financial support either for the activities themselves, the facilities and instrumentation required for research, or for research personnel or their mobility. The anticipated outcomes of research programmes are generally quantifiable and easily characterised (publications, qualifications, trained personnel, etc.).

In contrast, innovation programmes are concerned with a broader range of actors, locations and framework conditions which impact the innovation process: R&D forms but one limited (but nevertheless important) element within the innovation system and its role in innovation is subject to a diversity of factors such as technology and knowledge transfer flows, the capacity of actors to generate and absorb such flows, the conditions which influence the behaviour of those responsible for knowledge generation and uptake, etc. Consequently, by seeking to influence these factors, the outcomes of innovation programmes are inherently more complex and their effects and impacts may be more subtle or intangible, may take longer to manifest themselves and may be diffused across the innovation system. In short, innovation programmes are concerned with impacts and effects rather than outputs. Figure 4.2 illustrates the difference in the maturity of evaluation between research and innovation programmes.

It is the challenge of impact assessment that really differentiates innovation programme from research programme evaluation. This issue will be dealt with below, in Chapter 7.
Figure 4.1 Expert Views on Two Statements about the state of Evaluation Studies in this field

Note: as part of SMEIP, over 150 experts on research and innovation programme evaluation completed a survey concerning various issues arising. The chart above indicates that majority opinion disagrees with the idea that this is a mature area of work, and agrees with the idea that innovation policy evaluation is less well developed than that for research programmes.
4.11 How far can we use the methods and messages from Research Programme Evaluation in Innovation Programme Evaluation?

There is a very extensive literature on research evaluation – the field even has its own journal! There has been a great deal of discussion of practical and technical issues, and some attention to political and managerial ones. A substantial part of this is relevant to innovation programme evaluation, even though the sorts of instrument and impacts involved in research programmes are rather more straightforward (though they can be complex enough!). Many of the key issues and concepts in this Guide – the topic of additionality, for example – were first raised in research evaluation.

More general lessons also flow from the experience of research evaluation. Let us just mention two here. The first is that the evaluation process is more efficient and effective if it has been designed into the programme from the start – so that objectives are well-specified, data are generated and captured in a way that can feed the evaluation process, a strategy for the use of evaluation reports is in place, and so on.

Secondly, the importance of having well-informed sponsors and users of evaluations is critical. Even research programmes can have complex and multiple objectives, and users must be prepared to engage with the challenges that this creates for policy design – and for evaluation. Evaluations are liable to have to deal with different objectives using different data sources and methodologies, and are quite likely to need to clearly explain results in the context of a complex set of circumstances rather than tell a simple black-and-white story. Indeed, the story is liable to go beyond accounting for the programme itself, and will need to explicate various features of the research or innovation system of which it is a component. Users will have to be aware that they will need to engage with the understanding of these systems that the evaluation provides them with. Such an understanding is a vital input to policy formation and is more demanding than simply assessing whether a specific part of a programme has been effective to a particular extent. There is no “magic bullet”, i.e. a single methodology that is applicable to all programmes, and which yields information on programme performance without requiring (and hopefully contributing to) knowledge about the wider system in which the intervention operates. Such false expectations will limit the scope and utility of evaluation in innovation programmes, to a greater extent than in the case of research programmes.
Chapter 5 SCOPING THE EVALUATION OF INNOVATION PROGRAMMES

5.1 What is the evaluation for?

As argued earlier, evaluations may be conducted for a variety of purposes, under a range of rationales. But it can be helpful to bear in mind that, whichever specific purposes are top of the agenda any evaluation should first and foremost be a learning tool. It is undertaken in order to generate lessons for future practice, so that things may be done better in the future.

VINNOVA has suggested a fruitful way of thinking about the different sorts of contribution to learning that may be effected. The process of evaluation can thereby assist learning in three major ways:

♦ Operational learning: In this case, evaluation is used as a management feedback tool to improve the effectiveness, efficiency and quality of policy intervention by the organisation responsible for the implementation of the programme. Essentially, evaluation provides lessons on how organisations (ministries, agencies, etc.) can do things better, in terms of designing, managing and implementing programmes. Lessons may also be learned from the evaluation itself in order to improve the evaluation of future programmes.

♦ Policy feedback: Here evaluation is used in its “traditional” sense to determine the outcome and impacts of policy measures and programmes. This type of use is aimed at checking whether, and the extent to which, programmes have achieved their objectives. In addition to contributing to the design of future programmes, evaluation in this context also offers a method for policy makers to assess whether the assumptions they made about the identified bottlenecks and market or system failures which prompted the policy intervention in the first place were, in fact accurate.

♦ System impact: At this level, evaluations serve to improve the efficiency and effectiveness of national innovation systems through guiding the design and formulation of intervention policies and programmes. Evaluations are used to provide answers to broader level questions concerning the innovation system such as when certain interventions are appropriate, which complementary programmes should be used and when, what is the appropriate policy mix needed to achieve the desired effects, and so on.

(Source: adapted from VINNOVA, 2004)

If viewed in a very simplistic fashion, these three levels of approach encapsulate the evolution of a culture of evaluation, from the restricted management oriented view of programme administration, through evaluation as a justificatory process to a means to obtain a holistic and systemic view of innovation policy intervention.
Box 5.1 VINNOVA’s Strategic Vision

- **On the Policy Level**: Make evaluations and impact studies an integral part of the strategic policy planning and implementation process.

- **On the Programme Level**: Include plans for evaluation and impact assessments in the design, planning and implementation of innovation programmes.

- **On the Project Level**: Include relevant data collections implied by the policy and programme levels so that the requirements of monitoring, evaluation and impact assessments are fulfilled.

Source: VINNOVA 2004
5.2 What is the scope of an evaluation?

In the ‘traditional’ evaluation context, **scope** refers to a variety of parameters such as the type of programme to be evaluated, the level of the evaluation and the time-frame to be considered. When dealing with the evaluation of innovation support programmes, the systemic nature of innovation adds a further level of complexity to the evaluator’s decision-making process.

Our ultimate goal is to move away from the conduct of evaluations as isolated events, towards establishing their use as a feedback mechanism in a broader policy-learning process. Thus, the questions to be asked are not simply “has the programme worked” but “how has this programme contributed to our broader socio-economic goals”. If we wish to assess the systemic impacts of innovation support programmes, this demands that the scope of the evaluation must be extended to include a broader range of stakeholders, to examine wider impacts and effects, and that we must consider multiple evaluation methodologies, etc.

However, whilst a broad scope is desirable for policy-makers, pragmatic considerations, such as the timing of the evaluation within the policy cycle, the resources available for evaluation and the existence of appropriate methodologies will tend to constrain the scope to a more restricted set of prioritised objectives. Ideally, one must seek to balance the cost of the evaluation against the value of the information it gathers. Generally, we tend to encounter a law of diminishing returns; beyond a minimum threshold of essential information, as more resources are put into an evaluation, the utility of the acquired information tends to decrease. We consider evaluation costs in the next subsection. As for the utility of the information that is generated, there are several ways in this may be increased:

- Ideally, evaluations should generate data that is of relevance to the broader innovation programme and policy design process, rather than data which is highly context-specific.
- Methodologies should be selected which generate information and data which are easy to interpret and understand. Ideally, these results should also be easily translatable into policy responses.
- Sometimes a programme’s failure is obvious and does not necessitate an extensive evaluation. However, failure can be equally informative as success. Learning about the reasons for failure can help in the design of successful programmes in the future. Moreover, it is important to distinguish between the causes of failure: i.e. was it due to the quality or level of the inputs, the process, the results or to poorly identified gaps or bottlenecks in the innovation system or even to inaccurate assumptions about the nature or behaviour of the innovation system?
- If programmes share a common objective but adopt different policy approaches, then a series of evaluations will contribute more to the understanding of the policy mix and potential gaps, complementarities and mismatches.
SMART INNOVATION – Supporting the Monitoring and Evaluation of Innovation Programmes

♦ Ensure the dissemination of information is maximised in order that all actors in the innovation system benefit from the findings of the evaluation. In similar context, the institutional structure or programme design should be such that the capacity to make effective use of the results of the evaluation is maximised and that the evaluation results may be easily fed into the policy-making process.
5.3 What should be done before conducting an evaluation?

A key step in the development of an evaluation culture is to shift from the view of evaluations as one-off or isolated events and to recognise them as part of a cyclical process of policy design - policy implementation – policy learning. This entails a shift in the resources available for the evaluation, away from an emphasis solely on *ex post* evaluation, and towards a more balanced distribution of efforts across *ex ante* planning and design, monitoring and *ex post* evaluation.

A useful, if not essential, exercise at this stage is the design of a Programme Overview – also known as a programme model. One definition of a programme model is “a logically consistent description of the design of a programme as well as the expected impact on the effectiveness of the innovation system” (IKED/VINNOVA 2004).

Several forms of programme model exist, ranging from logic frameworks to ROAME statements (see below) and ‘balanced scorecards’. All effectively seek to map the systematic linkages between the policy problem (broad aims), the programme objectives, the activities or processes required to meet the objectives, and the anticipated results, outcomes, effects and longer-term impacts together with the necessary provisions for monitoring, evaluation and feedback into the policy making process.

![Stylised programme model](source: IKED/VINNOVA 2004)
5.4 Why is a Programme Overview Important?

A Programme Overview of some sort will typically have been developed when a programme is initially formulated and submitted for funding. It will at a minimum specify what instruments the programme is using, which sorts of actors are going to be involved, and what the aims are that should be accomplished. In practice, Programme Overviews vary from being extremely concise (just a few paragraphs or pages) to being very lengthy (which may be a matter of including considerable detail about legal requirements, or depth in explicating the intentions and principles behind the programme). The underlying rationale of a programme may be quite unclear – perhaps it is thought to be so obvious that it does not need spelling out, or perhaps there has been little reflection on the major policy goals involved – someone is convinced that policies are required in this area, perhaps having been persuaded by overseas experience or by influential lobbyists rather than through articulated and reasoned argument. Having an adequate Programme Overview, and an associated *ex ante* evaluation, should instil further rigour into this process, and the selection of programmes in general.

Likewise, it is common for the objectives embodied in a programme to be set out only in very general terms, with little quantification of targets (so it becomes difficult to judge whether what was aimed at and what has been achieved were reasonable and realistic). The concept of *verifiable objectives* has been introduced to indicate that the setting of objectives should be in terms which allow judgements to be made about the extent of achievements. While estimating the extent of achievements in advance is of course difficult (and we should avoid efforts at spurious precision), engaging in this can be an important learning process. (Not least, we may learn that different parties are talking about quite different things when they are specifying a “realistic” increase in innovation, a “sizeable” growth in employment, for example). The illustration on the facing page represents one way of thinking about key objectives.

Spelling out the objectives should also inform the case for the programme overall – why a policy intervention is believed to be necessary in this area. It should also indicate how various components of the programme are supposed to contribute to the whole, to the overall policy goals. A clear Programme Outline of this sort makes the evaluation process considerably easier – in practice, there have been far too many occasions where evaluators have had to retroactively construct what the Programme Overview should have been, in order to proceed with the evaluation.

While Programme Outlines can be prepared and presented in various ways, one formal approach that has received a great deal of attention is the ROAME framework, extensively used by the UK Department of Trade and Industry. Since this captures the main elements that need to be in an adequate Programme Overview quite concisely, we shall briefly explain how this works in a little more detail below.
Box 5.2 The SMART Approach to Best Output, Outcome and Indicator Statements

Try to frame these statements in terms that are:

S Specific

M Measurable (amount, speed, change, euros), and Meaningful (having “face validity”, obviously related to observable phenomena)

A Aggressive, yet Achievable

R Relevant (related to what you are trying to measure)

T Time bound (target is expected to be reached in a specified amount of time)

Source: based on, and developed from, Smith (1999)

A more detailed explanation of this is given in section 5.10
5.5 What is a ROAME statement?

ROAME is the acronym for an approach developed by the UK’s Department of Trade and Industry (DTI) in the 1980s, as a result of Treasury requirements to improve the management of programmes. Since its inception, it has been used widely across different government departments (Bradbury and Davies, 1996). A statement is to be drawn up at the outset – or even prior to the adoption - of a programme (not necessarily an Innovation Programme, the method is widely used for research programmes). The statement should comprise the following elements:

♦ **Rationale**
  This justifies the programme, setting out the policy goal(s) to which it is intended to contribute, and explaining why the programme is necessary (in terms of market or system failures).

♦ **Objectives**
  This states specific (and ideally measurable) objectives that should be achieved – an “operationalised” statement of contributions to the overall policy goal. The statement of objectives should include specification of performance indicators that provide feedback on the success of the programme.

♦ **Appraisal**
  This sets out the choice of activities that are selected in order to achieve the programme objectives, explaining why these particular activities were chosen and justifying their selection.

♦ **Monitoring**
  This explicates the procedures to be used routinely during the course of the programme to check that its progress is proceeding as planned. Monitoring requires flows of information related to management (how far activities are taking place as planned) and results (what is progress in achieving the objectives, e.g. as specified by the performance indicators).

♦ **Evaluation**
  As discussed in this Guide, this involves examining the efficiency with which the programme has been conducted, and its effectiveness in terms of achieving the intended objectives. An attempt is also made to account for (positive and negative) wider effects of the policy or side-effects that were not envisaged initially. It may also examine issues to do with programme rationale and objectives, and with the unexpected consequences of, and lessons learned during, the programme.

Sometimes the acronym has been extended to ROAME-F, with the additional letter connoting:

♦ **Feedback**
  This is a statement of how the evaluation results are intended to inform the strategy and design of subsequent programmes: for instance, procedures for publishing, disseminating, and gaining responses to the evaluation.
5.6 When should ROAME be used?

Procedures such as ROAME have the considerable virtue of making programme designers be very explicit about what they are doing, and setting out why this course of action is justified and has been selected against alternatives. By making the overall aims of the programme clear and concrete, they make the task of evaluation cognisably easier, since the actual achievements of the programme can be benchmarked against those that were intended. The specification of performance indicators should mean that much information that is highly relevant to the evaluation will have been generated in the course of the programme (and captured through an efficient monitoring system). There should thus be fewer requirements for new data to be generated, and more chance that information that might otherwise have been lost during the course of the projects will have been retained (although the evaluators may wish to define additional indicators, if the experience of the programme indicated to their judgement that the initial performance indicators are inadequate in one or other way).

ROAME sets out various elements that, if specified at the outset of a programme, will allow for better clarity about what the programme is intended to do; will define many of the procedures to be used in the course of the programme; and will make the task of evaluation much simpler. It will be apparent that following a ROAME-type approach implies that ex ante evaluation will be undertaken before the actual beginning of a programme (during its design, in fact), and that instructions will be in place setting out (in broad terms) how ex post evaluation is to be employed at the end of a programme. It thus builds evaluation into programme design. This facilitates the end-of-programme evaluation, and makes it more likely that it will be successful in terms of achieving high quality and impact.

But ROAME as a management approach does more than contribute to evaluation. It should prompt decision-makers to think systematically about what they are doing when they launch a programme, and to institute procedures that can improve programme management. Before there is any feedback of evaluation results, there should be feedback of information about how programme activities are proceeding. This will allow for corrective actions to be undertaken should problems arise – indeed, it should allow for the more rapid detection of problems. It also permits better communication with the stakeholders involved in the programme – the statement will provide clear information to all parties about what the programme is intended to do, how specific activities are contributing to this, and what management procedures will be put in place for its implementation.
5.7 What is the “Business Case” Approach?

Very recently the use of ROAME-F in the UK Department of Trade and Industry has been superseded by the use of “business case documents”. Any (internal) proposals for innovation and business support measures and programmes (and which cost in excess of £10 million/€14 million) are supported by a business case. Business cases must be approved by the Investment Committee, and the Secretary of State for Trade and Industry, before funds are released for their realisation. As under RAOME-F, during the lifetime of a product or programme, it will be subject to ongoing monitoring and analysis, and periodic impact evaluation.

The business case documents take the place of “ROAME” statements. They continue to set out the rationale and objectives for a new product, with particular attention to the supporting evidence, and they also provide a description of the proposed form of intervention and how it will be delivered. However, the business cases also include a number of new features that provide useful information for evaluators and help to support improved monitoring and evaluation:

- Option appraisal and cost benefit analyses: A record is provided of any alternative policies that have been considered, and also of key assumptions about the expected costs and benefits associated with the proposed option;
- Logic model: This outlines the inputs and activities, and their anticipated outputs and outcomes. The logic model has a key role in identifying monitoring and evaluation measures and indicators;
- Balanced scorecard: The balanced scorecard translates the product’s logic model into a summary set of objectives, and associated monitoring measures which will be used to track progress against these objectives. The scorecard focuses on measures for which meaningful monitoring data can be obtained within quarterly and annual reporting periods, and takes account of the time frame over which the product is expected to generate economic outcomes.

(Malik, K and Cunningham, P, 2005)

In addition to balanced scorecard reporting, the DTI also makes use of monitoring surveys of business users, intermediaries and delivery partners, data-linking and econometric studies and impact evaluation.
5.8 What is the role of the Programme Overview?

From the evaluation perspective, spelling out the logic of a programme is a key component of the policy design - it encourages the participants to define clear targets and objectives at the outset of the policy process. Likewise, by mapping the innovation system into which the policy will be introduced, a deeper understanding is gained of the innovation process and of the desired and potential results, effects, outcomes and impacts of the policy itself.

This is a significant step away from the so-called “black box” perspective in which evaluations seek to establish correlations between inputs (resources, policies, etc.) and outputs towards a more transparent understanding of the actual consequences of the introduction of a policy.

The construction of a programme model is in itself a highly valuable exercise as it forces programme managers and policy makers to form a holistic overview of the programme or other policy intervention and to question the broader systemic aspects of its introduction, for example through linking process and action to rationale and by identifying the full range of stakeholders. Once the stakeholders have been identified, it is equally important to attempt to engage them in the evaluation design process. The VINNOVA handbook suggests that one way to involve the actors in the planning of the evaluation process is to make the formulation of a programme model an obligatory part of all applications for public grants. This is akin to the approach adopted by the UK Department of Trade and Industry’s business cases model (see discussion above section 5.7).

The construction of a programme model forces the actors to state clearly the programme objectives and to identify the causal links between activities and expected effects. This focus on targets and objectives prompts the consideration of evaluation questions at the beginning of the policy-making process. The early involvement of actors also increases their influence on the selection of indicators and appropriate methods for measuring impact. In turn, this changes the relationship between the evaluators and those being evaluated and enhances the move away from evaluation as a control measure towards the recognition of evaluation as a learning process. In some cases the evaluators can act like process consultants, bringing energy and focus into the programme process. (This may be particularly true for real-time evaluation.)

Lastly, the construction of a programme model also serves to present a concise, one or two page visualisation of the policy process to other policy makers, politicians, stakeholders and other interested parties.
5.9 What is the Balanced Scorecard Approach?

The UK’s Department of Trade and Industry (DTI) have had a culture of doing evaluation from the 1980s and that’s evolved and the latest stage in the evolution is that for the business support products the DTI are now using the ‘Balanced Scorecard’ approach. The balanced scorecard was introduced in 1992 by Robert Kaplan and David Norton, as a set of measures that allow for a holistic, integrated view of business performance. The scorecard was originally created to supplement “traditional financial measures with criteria that measured performance from three additional perspectives — those of customers, internal business processes, and learning and growth” (see Kaplan and Norton 1996).

Key to considering the use of a balanced scorecard methodology for any types of innovation programme are the steps that link the larger goals of the programme to more specific problems that have to be solved, decisions to be made, and resource allocation choices that present themselves. While the balanced scorecard cannot guarantee a recipe for correct decisions, it provides an integrated perspective on goals, targets, and measures of progress. It ties together information from a variety of perspectives so that trade-offs can be weighed.
Box 5.3 The Balanced Scorecard Approach in the UK

The DTI (in the UK) have recently started to use the balanced scorecard approach. An evaluation team leader in the DTI, who was interviewed for this project summarises this approach as follows:

“The balanced score card is a methodology that looks at a particular programme from the following four perspectives:

• Performance
• Relationship within the programme between the beneficiaries and any intermediaries
• Individuals who are providing support
• Processes used

For each of these 4 perspectives, the Balanced score card sets measures in both the intermediate and long term, and can measure inputs, outputs, impacts and outcomes. The score card is designed to identify decided testable and measurable objectives, with an aim to focus on relatively few, but key measures. Data from these measures is gathered regularly – on a quarterly basis and also at an annual point. The data is then fed into a software tool which underlies the methodology. It is then possible to interrogate the database, enabling evaluators/researchers to produce reports and record views, on how any given programme is doing at a particular time”.

Evaluation Expert (UK)
5.10 How can we make statements of objectives more useful for evaluation and monitoring?

A major consideration in selecting objectives is that their achievement should be at least verifiable and, ideally, measurable. That is to say, it should be clear that the selected objectives have or have not been achieved and also the extent to which they have been achieved should be quantifiable or determinable in a robust qualitative manner.

The setting of verifiable or measurable objectives is a useful task as it:

- Clarifies and makes explicit the link between the programme or activity-level objectives and higher level policy aims and objectives.
- Creates a direct link between the problem to be addressed and the analysis of what needs to be done.
- Can help to provide a common understanding of what are the important aims and activities which can assist in the implementation of the programme.
- Provides a clear basis for the definition of indicators which may be used to measure progress and assess achievement.
- Lays the basis for the evaluation of the programme and assists the evaluators in determining the degree of success the programme has achieved.

A set of desirable properties for objectives is provided by the European Commission (2002). This recommends that, if objectives are to be useful they must be:

- **Specific**: Objectives should be precise and concrete enough to avoid being open to varying interpretation.
- **Measurable**: Objectives should refer to a desired future state (as compared to the baseline situation). So that it is possible at some later date to see whether the objective has been achieved or not.
- **Accepted**: If objectives and target levels are intended to influence behaviour, they must be accepted, understood and interpreted similarly by all of those who are expected to take responsibility for achieving them.
- **Realistic**: Objectives and target levels should be ambitious – setting an objective that only reflects the current level of achievement is not useful – but they should also be realistic so that those responsible see them as meaningful.
♦ **Time-dependent**: Objectives and target levels remain vague if they are not related to a fixed date or time period. In this context, the setting of milestones for the partial achievement of objectives is useful.\(^8\)

Other considerations for the setting of objectives are that they should be **limited in number** and should address only key elements of the programme. It is likely that when a large number of objectives are being talked about, the programme activities have not been properly clarified, the programme remains too unfocused (and will quite possibly contribute minimally to many goals rather than substantially to any of them). Similarly, a large number of objectives will require the identification of an extensive range of indicators and it will be difficult to measure their achievement effectively, particularly if time, personnel and financial resources are limited.

---


5.11 What criteria should be used for the evaluation?

The criteria for the evaluation of an innovation support measure depends upon the specific nature of the intervention. However, in general terms innovation programmes are likely to exhibit a number of characteristics in the effects they are designed to have. These may encompass:

♦ Effects on the firm or other participant
♦ Effects on industry more broadly
♦ Collaboration effects
♦ Effects on the broader portfolio of innovation support projects/programmes
♦ Synergy effects
♦ Consistency and sustainability effects

Whatever the specifics of an evaluation approach, the ultimate goal is to measure the effectiveness of the intervention in the context of its real impacts. Thus, innovation support measures should be evaluated on the basis of a set of comprehensive criteria. These may comprise:

♦ Appropriateness: do the programmes address an objective that can be clearly related to policy rationale in terms of market, policy, system or government failure?
♦ Superiority: Are the programmes more effective than other sets of instruments that might have achieved the same goals?
♦ Systemic effectiveness: How does the policy interact with other policies? Is there efficiency or sub-optimality in the broader set-up?
♦ Efficiency: Are the programmes cost-effective in achieving their specific objectives?
♦ Adaptive Effectiveness: To what extent do the results from evaluations feed back into policy design and implementation? Does policy design ensure a sufficient degree of flexibility enabling it to respond to the need for change?
5.12 What about the financial costs of evaluation?

The costs of an evaluation should obviously not be disproportionate to the costs of the programme it is intended to evaluate. A rough rule of thumb has been that evaluation costs are 1-5% of overall programme costs. This is notional, because some of the work of evaluation should be embedded into the routines of the programme, for example the data that are produced for routine monitoring.

There are various ways in which the costs of evaluations can be reduced, among them:

- Having a clear hierarchy of information requirements in order that the collection of data is prioritised in strict accordance to its utility.

- Developing sets of common indicators, which may be collected routinely by participants in a number of related or similar programmes. This process also tests the feasibility of using particular indicators and increases the effectiveness of their use. The possibility of using sets of indicators to compare related programmes forms another benefit.

- Indicators, clearly linked to programme objectives, should be defined prior to the initiation of programmes and should be collected, where appropriate, from the start of a programme. This not only permits ongoing monitoring and enhances programme management but also builds a database of consistent indicators which may be used by evaluators (either internal or external) without recourse to an expensive post hoc data gathering exercise.

- Prioritising the use of methodologies that have a minimal impact on those evaluated and maximise the rate of information return (telephone surveys over face-to-face interviews, email surveys rather than mail surveys).

- Gaining the trust and compliance of all stakeholders, through demonstration of the utility of the evaluation process, in order to minimise implicit or explicit resistance to the evaluation process.

- Ensuring that evaluators have the required competencies and exhibit a track record in the range of methodologies required for the evaluation.

- If evaluation is explicitly acknowledged to be part of a learning process, then some of the evaluation costs can be seen as training costs, and to the extent that these reduce costs for ordinary training within the organisation, some funds may be acquired from this source.
5.13 What should the timing of an Ex Ante evaluation be?

As already discussed, an evaluation can take place at three possible instances – before the programme is launched (ex ante assessment), whilst it is in progress (real-time, ongoing evaluation or monitoring) or after the programme has ended (ex post). When to undertake an evaluation has always been a major issue for programme managers – too early and the results may not be evident, whilst longer-term impacts and effects will certainly not be. On the other hand, programme and policy makers need information on the success or otherwise of a programme as early as possible in order to maximise its effectiveness and efficiency.

Clearly it would be impossible to assess the actual results and impacts of an innovation support programme prior to its launch. However, as we are advocating that evaluation should form an embedded component of programme management and the policy formulation process, the anticipated effects and outcomes of the programme should already be clearly understood as they provide the rationale for the programme itself. Indeed, the design of an innovation policy measure or programme does not (and should not) take place without a thorough ex ante appraisal in which the (market or system) failures are clearly identified and the means to address these failures are formulated. Nevertheless, this does not mean that a programme may not be evaluated prior to its initiation. Two options are available:

♦ The expected outcomes and impact of the intended programme may be modelled, in the context of the set of implicit assumptions about the innovation system in which it is intended to intervene. Although relatively sophisticated, this approach can contribute to building the rationale for the programme and, from a comparison of the expected result with the actual outcomes, may provide useful information on the validity of the underlying assumptions about the innovation system and also help to identify any unintended spill-overs and externalities.

♦ The programme may be launched, on a reduced scale, as a pilot which may be subject to an evaluation prior to rolling out a full-scale version. The results of the evaluation may then be used to inform any necessary modification to the full-scale programme, thereby reducing the risk of failure or the inefficient use of resources.
5.14 What should the timing of a Real-Time Evaluation be?

It is often desirable, particularly with programmes that are not expected to generate results and impacts for a considerable period of time, to gauge, at least partially, some aspects of their performance. There are several reasons for this: badly performing programmes may be modified or even terminated, the framework conditions or environment may alter significantly from that which existed when the programme was initiated, or the investment in long-term programmes demands that continued justification for their prolongation exists.

Monitoring and real-time evaluation offer two methods to assess the performance of programmes whilst they are still operating. Strictly speaking, monitoring represents a relatively low-level, low resource-intensive process of data and information gathering which enables programme management to make periodic assessments of progress. The process requires the definition of a number of clearly specified types of information, data or indicators which may be routinely captured, without intensive effort and collated in a programme database. As it results in the compilation of a useful set of background information on programme processes and outputs, monitoring has the additional benefit that can reduce the burden of data gathering often entailed in larger-scale, ‘one-off’ evaluations.

Real-time evaluation is a term used to distinguish ex post evaluations from those conducted whilst the programme is still operating. It differs from monitoring in that it attempts to provide a deeper understanding of the results and outcomes of a programme and looks beyond routinely collected, monitoring, information (although this may feed into the real-time evaluation) to other types of indicators. It also relies on a broader range of methodologies and may involve a wider set of stakeholders. The objective of real-time evaluation is to provide an input into the management of the programme through obtaining a broader picture of the programme’s performance. It can be used to influence decision making on the continuation of the programme and on any required modifications and adjustments.
5.15 What should the timing of Periodic Evaluations be?

Periodic evaluations are often used to assess the performance of long-term innovation support programmes. They are typically planned as a series of predetermined evaluations – i.e. it is known that each one will be conducted after a specific time period. They may be conducted on a regular basis (for example, every five years) or timed to coincide with events in the wider policy cycle (for example, large-scale funding reviews, budgetary cycles, etc.). They share many of the features of *ex post* evaluations in terms of the methodologies employed and the fact that they are essentially backward looking and focused on historical evidence, although the programme is still on-going. Periodic evaluations offer a number of advantages:

♦ Through using standardised indicators, or indicators used in preceding programme evaluations, they produce valuable comparable time-series data on the long-term performance of the programme which may provide insights against the context of variations in the wider innovation system framework and economic conditions.

♦ Alternatively, each evaluation may focus on different aspects of programme performance, perhaps through the use of different methodologies, in order to provide a broader picture of the programme’s outcomes and impacts. The results of earlier periodic evaluations may also generate additional or unexpected lines of enquiry of interest to policymakers which can be followed by subsequent evaluations.

♦ Unless the programme is designed to respond to a short-term problem, gap or bottleneck, it is likely that an innovation support programme will have a substantial lifetime, particularly if it forms a component of a mix of policies. In this case, information to aid programme management and higher level policy decision-making information will be required earlier rather than later. As noted in the above point, tailoring the focus of the periodic evaluations, from process issues in the early life of the programme towards the measurement of outcomes and impacts in the later stages of the programme’s life may be a useful practice.
5.16 What should the timing of an Ex Post evaluation be?

As their name suggests, these evaluations are conducted after a ‘fixed-term’ programme has been completed, or has reached a specific phase of its lifetime. Generally, ex post evaluations tend to be one-off in nature and provide information on the degree of success that a programme has achieved. Of course, despite the fact that the programme may be a ‘one-off’ policy intervention, as in all evaluations it is critical that the outcomes and lessons from the evaluation provide an input into policy making and the formulation and design of any potentially similar future programmes.

The issue of timing is perhaps at its most sensitive in the context of ex post evaluations, although instances are known where an immediate ex post evaluation has focussed on process and outcomes-related issues, and a later evaluation has focused on longer-term effects and impacts. Of course, a number of evaluations of different types may also be conducted on a programme; one example is provided by the Norwegian “Evaluation of User-friendly Research in Norway” (See Box 5.4).
### Box 5.4 The modular economic evaluation system

<table>
<thead>
<tr>
<th>Assessments</th>
<th>Time of evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ex-post evaluation</strong></td>
<td>Just after programme termination</td>
</tr>
<tr>
<td>Short-run economic and external effects</td>
<td>Just after programme termination</td>
</tr>
<tr>
<td><strong>Long-term evaluation</strong></td>
<td>About ten years after project start-up</td>
</tr>
<tr>
<td>Economic impacts and external effects</td>
<td>About ten years after project start-up</td>
</tr>
<tr>
<td><strong>Infrastructure evaluation</strong></td>
<td>Just after, then ten years later</td>
</tr>
<tr>
<td>Network and external effects</td>
<td>Just after, then ten years later</td>
</tr>
<tr>
<td><strong>Econometric evaluation</strong></td>
<td>Time series</td>
</tr>
<tr>
<td>Time series analysis combining company and project information to estimate rate of return</td>
<td>Time series</td>
</tr>
<tr>
<td><strong>Ex-ante evaluation</strong></td>
<td>Year after start-up</td>
</tr>
<tr>
<td>Start-up expectations of economic and external effects</td>
<td>Year after start-up</td>
</tr>
</tbody>
</table>

5.17 How can the objectives of individual evaluations be specified?

One of the ‘golden rules’ in evaluation is to define clearly the policy objectives that are to be addressed by the programme. This is why so much stress has been placed upon the Programme Overview in the discussion above.

In the past, the question, posed by evaluators, “what were the objectives of the programme?” has often received no answer or at best a vague set of aims. In developing a culture of evaluation, it is desirable that objectives are set during the programme formulation process; indeed, the linkage between policy goals and programme achievements is a key element of the programme design logic. Objectives are best defined as the expected or desired effects of the programme on the situation that it has been designed to influence. As such they embody an idea of a change from a baseline situation and are clearly linked to the problem, bottleneck or gap that is to be addressed by the programme. Obviously, once objectives have been specified, it is possible to define indicators which may be used to measure or monitor the results of the programme.

However, before looking at the specific objectives of an innovation programme, some discussion of the actual level of the objectives is necessary. It is clear that a well-designed programme should impact a number of policy levels; for example, a political desire to increase employment in high-tech industries may be achieved through a programme aimed at the creation of new-technology based firms - hence a range of objectives may also be anticipated. These can be categorised as:

- Programme or operational objectives: the set of desired outputs or deliverables from the implementation of the programme. Generally, the achievement of such objectives may be readily verified.
- Broader policy objectives: these include goals such as increased employment, economic growth, regional growth, export promotion, enhanced competitiveness, etc. The achievement of the programme objectives should influence these broader objectives, at least partially.
- Impact on framework conditions: innovation support programmes may (and increasingly do) have indirect effects on components of the environment for innovation such as the tax system, regulatory frameworks, competition policies, IP systems, etc.).

The rationale for the innovation programme or intervention should clearly delineate the various objectives that are desired.
5.18 What types of indicators should be selected?

Indicators are the measurable outcomes which show whether the programme objectives are being achieved.

General guidelines for the selection of indicators are that they should be:

♦ **Relevant**: There should be a clear intuitive link between the indicator and the objective that is to be achieved. Stakeholders will gain a better picture of the programme’s performance if the indicators reflect the output or impact end of the spectrum rather than being input-oriented. Output indicators may also be combined with input indicators to provide an assessment of the relative performance of the programme, for example, with respect to other programmes.

♦ **Bounded and comprehensive**: The information provided by the indicators should focus on a small number of the most significant indicators and it should also be comprehensive, i.e. this set of selected indicators should cover all the main aspects of programme performance.

♦ **Accepted**: Indicators should be discussed with all programme stakeholders and interested parties to ensure that that agreement has been reached on their interpretation and acceptance.

♦ **Credible**: Are the indicators easy to interpret? Are they credible for reporting purposes? Are they unambiguous to a variety of stakeholders?

♦ **Easy**: Are the data accessible and readily obtained. Can they be easily measured without excessive or disproportionate effort (i.e. does their value as information justify the costs of their capture?). Ideally, indicators should be easily and efficiently captured through monitoring during the life of the programme rather than through more expensive *post hoc* exercises. In addition, indicators should also be capable of independent verification.

♦ **Reliable and Robust**: The information must exhibit an appropriate level of accuracy and dependability against the purposes for which it has been collected. Are the indicators proof against manipulation by those responsible for their production? Could the setting of such indicators lead to unwanted distortion of the behaviour of those responsible?

♦ **Consistent and comparable**: In an ideal situation, indicators should be consistent from year to year although they should also be refined and modified when necessary to reflect important contextual changes. Where possible, indicators should also be comparable so that like may be compared with like, either within or between programmes.  

---

5.19 What are the advantages of a quantitative focus?

Some measure of quantification is necessary in programme planning and evaluation, if only because funds are being allocated, and money is inherently a quantification of specific efforts and resources.

Quantitative indicators have several advantages:

♦ Comparability across projects: they enable comparisons to be made with the outcomes of several projects or programmes which share the same broad attributes.

♦ Comparability through time: Similarly, they enable comparisons between different time frames and permit longitudinal analyses.

♦ They are easily presented in a variety of formats and hence are more easily interpreted by both internal and external audiences.

♦ They are inherently attractive to policy makers and politicians as they are (apparently) free of contextual details and associated caveats about their use. This loss of context and explanatory detail unfortunately opens them to the possibility of misinterpretation.
5.20 What are the advantages and disadvantages of a quantitative focus?

Some measure of quantification is necessary in programme planning and evaluation, if only because funds are being allocated, and money is inherently a quantification of specific efforts and resources. The questions start to arise where we come to quantification of the (intended or achieved) results of evaluations.\(^{11}\) Here we encounter several sorts of issues:

- **Practical problems** – some topics are inherently difficult to quantify. Even when this is not evidently the case, some topics will have had quantitative indicators (even regular statistics) established for them, while others will not have been addressed in this way (and thus the evaluators may have to construct new indicators).

- **Psychological problems** – it is apparent that quantitative data are particularly appealing, especially to people with certain cognitive styles. Often it is the tables in a report that people home in on, rather than the text that accompanies the table. Several problems arise. (1) Overemphasising those things that can be or have been quantified, and neglecting or even dismissing (e.g. as “unscientific”) those that are better expressed qualitatively. (2) Failing to recognise that even the best quantitative indicators will be only partial ones – like all indicators that capture only some dimensions of the issue we are trying to illuminate – and slight variations in indicator design may mean that we are capturing quite different features of something, even though this is not immediately apparent (and the indicator may still be referred to by the same name). (3) More mundanely, neglecting the qualifications that often accompany quantitative data, for instance about the problems of drawing inferences from samples of a given size, about the possible error margins for estimates, about cases of missing data, and the like.

- **Cultural problems** – there seem to be variations in the extent to which quantitative approaches are stressed or distrusted in different settings. This may have something to do with the educational systems of different countries (or the processes of recruitment of civil servants), but it does mean that in practice we cannot always expect to directly copy a set of procedures developed in one environment, in another. Some environments are highly receptive to quantitative information, in some cases to the extent that there is acceptance of “spurious precision”. Some environments are much less so, with participants suggesting that numbers can be made to say anything, and demanding narratives and stories as a way of apprehending the underlying situation in depth. (Similarly, use of figures and illustrations is something that works much better in some circumstances than in others.)

However, there are very clear virtues in using quantitative data where these are available and actually appropriate to the problem at hand. Such data can provide a background to a rich qualitative analysis; they can be more readily

\(^{11}\) See OECD (1997b) for some discussion of these issues.
and transparently processed (summed, averaged, cross-tabulated, etc.) and transformed into visualisations such as charts. It is possible to apply statistical tools to test for significance,\(^\text{12}\) to control for the effects of different variables (e.g. whether the programme impacts vary across firms of different sizes, in different sectors).

\(^{12}\) But note that significance texts are widely used where they are not technically justified – they were designed to provide evidence as to the generalisability of results from samples to a wider universe of cases, and have strict assumptions about the nature of the sample, the parametric properties of the data, and so on. Often tests are in practice used with little concern for these assumptions, and with no real effort to specify what universe the results might be generalising to – the text is used as a signifier of “statistical importance” of the results. (Which is not the same as practical importance.) This practice may be defensible, but it needs to be understood that this is what is being done, not just accepted as unproblematic practice.
Chapter 6 CHOOSING AND USING METHODOLOGIES FOR EVALUATION

6.1 Why is Methodology Important?

Evaluation methods are tools and techniques that are applied in order to obtain answers to specific questions. The methods used in evaluation research are in the main ones that are used widely in other social and business research – just a few approaches (such as cost-benefit analysis and impact assessment) are really distinctive.

Methodology involves understanding the principles on which particular methods are based, and thus being able to understand the relevance of various methods to answering the questions which we have. We will often find that any given method will give us only a partial answer to the main questions that we have. This is because our questions are typically rather complicated ones, which cannot be answered in any simple way.

In evaluation studies, we will typically use a number of specific methods. For example, it is common for a study to include some interviews alongside a broader survey, and some use of existing (secondary) data. Alongside this there may be case studies of particular firms or projects, and a range of other approaches. There have been accounts of “mixed method” evaluations as if these were rather novel, but in the innovation programme evaluation field, at least, they are the norm.

This chapter focuses on the methods used in evaluation, and provides a broad introduction to methodology and methods, relating these to different types of evaluation and evaluation question. It begins with discussion of how methodology relates to the purpose of evaluation.
<table>
<thead>
<tr>
<th>Box 6.1 A Selection of Relevant Guides to Evaluation and Research Methodology</th>
</tr>
</thead>
</table>
This is a very readable hypertext guide to social research methods, covering a huge range of topics in a lucid manner. Its sections on reliability and validity of indicators are particularly recommended. A paper copy and expanded version is also available, from Atomic Dog Publishing at [http://www.atomicdog.com/trochim](http://www.atomicdog.com/trochim) |
| **Carter McNamara** (n.d.) **Basic Business Research Methods**, available online at: [http://www.managementhelp.org/research/research.htm](http://www.managementhelp.org/research/research.htm)  
Wide-ranging review of research methods, with good advice on planning a study. |
Detailed account of the wide range of methods used in US research programme evaluation, with some overview and synthesis concerning lessons for evaluation research more generally. |
This is the “ePUB RTD Evaluation Toolbox” prepared by the Joanneum Research Institute of Technology and Regional Policy (Vienna) and the Institute for Prospective Technology Studies (Seville), with contributions from many evaluation teams. As well as introducing the range of techniques in use, this report examines some of the wider issues around use of evaluation in policymaking. An abridged version is available as InTeReg Research report 02-2002 from Joanneum Institute. |
Useful overview of issues in evaluation studies, especially those concerned with the need to use both qualitative and quantitative methods in the course of a study. |
Though the sorts of programme being evaluated here are far from innovation programmes, this is a very clear introduction to the range of methods used in evaluation practice, with particular emphasis on using evaluation for empowerment of people involved in programmes. Activities are presented in terms of logical sequences of steps, in a manner that could be emulated in innovation programme evaluation. The website indicates that “empowerment evaluation” approaches are now being downplayed in Americorps. |
Though the programmes for which this toolbox was prepared are far from innovation programmes, this is both a readable guide to practice and a good introduction to the philosophy of participatory monitoring and evaluation. |
This guide should help practitioners of evaluation in the EU commission, who need to plan, co-ordinate, carry out, or utilise evaluations. It will also assist those with a responsibility for organising the evaluation structure within the services. |
6.2 How does the Evaluation Mission relate to Methodology?

The key issue for methodology is: how can we best access and use evidence that goes at least some way toward answering the question that we are posing? There are several points raised by this.

First, “best”. The question is not about perfect intelligence: that can never be achieved. The “best” evidence will be the most adequate evidence that can be obtained within the limits of available resources — including the time constraints (imperfect knowledge when it is needed can be a great deal better than more complete knowledge that comes too late). Often we need to use more than one method to provide an adequate answer to our starting question. Triangulation is the term used to describe application of several methods and observations to address the same question — if the results coincide, this validates them, if they diverge then it demonstrates that we have further work to do to undertake to understand this divergence.

Second, “access”. The evidence may already be available — we talk of secondary data when we are making use of, for example, statistics already produced for a different purpose. Alternatively, we may need to go out and generate our own primary data.

Third, “at least some way toward answering”. This reflects the reality that often the full answer will be hard to ascertain. In the case of evaluating the effectiveness of a new drug or medical treatment it may be relatively easy to assess impacts on the condition that it is meant to treat (though side-effects may be harder to determine). In the case of innovation programmes, many of the effects will be long-term, and the ways we have of knowing whether the effects are taking place may be partial and limited. This is why it is so important to be explicit about what question we are posing. The following section deals with the critical issue of operational definitions.

The literature on evaluation methodology provides us with much insight into the applicability of different methods, and we draw on this (as well as on the experience of the people we have interviewed and surveyed) in the remainder of this chapter.
6.3 What is the significance of Operational Definitions?

Everyday language uses many terms in rather loose ways. While this means that we are not forever engaged in pedantic or technical discussions, there are drawbacks. Often, we are not entirely sure what each of us is referring to when we talk about things like “intelligence”, “competitiveness”, or “economic performance”. This even – some would say especially – applies to political debate. It becomes a real problem when we are discussing specific policies, however, where there may be misunderstandings that can have big consequences for the real world.

An “operational definition” is intended to add precision, by providing a definition of (elements of) a concept in terms of how we are measuring them. The advantages of operational definitions are straightforward:

♦ They ensure that we are talking about the same thing
♦ They specify methods used to assess the elements of the concept that are being studied, and this allows for debate about the content of the concept
♦ They mean that work should be replicable, in that it is possible to follow the same methods to obtain further results – perhaps to check the original study, perhaps to update the information

The main disadvantages of operational definitions are a matter of how they are used:

♦ There is often a tendency to assume that the readily measurable features of a concept are the most important, or even the only, features of the concept. (Thus, famously, “intelligence” becomes what intelligence tests measure, not the richer everyday use of the term.)
♦ There is often overemphasis on data that can readily be displayed in a quantitative form (which many operational definitions allow), as opposed to more qualitative results.
6.4 Why is there no single Methodology for evaluation?

Usually measurement will only tell us about certain aspects of the issue in question. For instance, it may seem a simple matter to determine whether a car or computer is value for money. But consider how many aspects there are to the performance of a car or computer (let alone aspects such as the design and “feel” of the product). Similarly, policymakers may ask a simple-sounding but complicated question such as “what are the economic benefits of this investment”. The evaluators will have to specify a researchable question in terms of operational measures, such as “how did this programme influence the performance of X [a particular set of economic actors] in terms of Y [a particular set of parameters of activity and outputs]?”

In evaluation studies, the impacts and outcomes of interventions are usually highly complicated, with different sorts of target experiencing impacts on several dimensions. Some of the impacts may be captured in official statistics, company reports, and other documents; some may require direct investigation. Some of them may be a matter of counting observable phenomena, some may be more a matter of subjective judgement. Some may be a matter of individual experience, some of organisational behaviour, and some may even cross organisational boundaries (e.g. networks).

The diversity of methods available for performing an evaluation is an acknowledgement of the multiple dimensions in which the impacts of policy intervention might manifest themselves. For this reason, no single best evaluation methodology exists, for all purposes and studies. Each methodology will be more suitable for analysis of particular dimensions of impacts. In general, an evaluation study will require a combination of various evaluation methods. Thus, for instance, different methods may be used at different levels of data aggregation, or to capture immediate and longer-term impacts. The use of more than one method has advantages in that it allows for cross-checking the robustness of conclusions about the observed effects of the intervention – i.e. it permits triangulation.
6.5 What are the Main Methods used in Evaluation?

While the precise impacts addressed may be different ones, the general methods applied in research evaluation studies are typically applicable to innovation programme evaluations. As already mentioned, most of these methods are also shared with a great deal of other social and business research.

Some “methods” are very specific techniques, while others are more general approaches to answering questions. Thus questionnaire surveys are specific techniques (and within these there are other specific techniques being employed, for example in sampling, in the design of measurement scales and questions, etc.). Case studies, in contrast, for more of an approach to developing evidence and arguments, that can utilise very different sorts of technique within them – surveys, interviews, workshops, and so on.

In this chapter we will focus on some specific techniques and approaches that are widely used, grouping these according to the functions that they play, the specific information that they contribute:

- Methods for accessing and generating data – techniques of data production that are used in primary and secondary research. Here, the sorts of indicator that we are able to use is a major issue.
- Methods for structuring and exploring interventions – ranging from “traditional” approaches using experiments and control groups to more action and participatory approaches.
- Methods for analysis of data – ways of processing and drawing conclusions from statistics and qualitative material, and more elaborate modelling and simulation approaches.
- Methods for drawing conclusions, including impact assessments and the like.

As well as providing a panoramic view of the various techniques used here, we will also discuss certain thematic approaches – the use of multiple methods, the case study approach, Foresight, and the like.

There are various guides to social research methodology, to evaluation methodology, and to research evaluation methodology, which we can draw upon here. The area closest to that dealt with in this Guide is that of R&D (or RTD) Evaluation, and thus sources such as the ePUB RTD Evaluation Toolbox (2002) are particularly useful – see illustration.
**Box 6.2 The “ePUB RTD Evaluation Toolbox” (2002)**

This “toolbox” of evaluation methods provides a good starting point for finding out about the main instruments applied in science and technology policy evaluation and the current methods and practices used in the evaluation of public RTD policies. Based on the efforts of a network of evaluation experts, it discusses some of the political issues concerning methodologies, as well as the more practical ones.

Most ePUB Toolbox methods focus on the capture of data relevant to RTD programmes. At the micro-econometric level, this includes methods for evaluating the existence of an additionality effect in public support on private R&D and estimating the private rate of return on R&D. More macroeconomic methods are better fitted to capture the generated R&D spillovers and longer-term and wider effects of the policy intervention on productivity and economic welfare. Specific tools discussed include cost-benefit analysis, network analysis, case studies, and the like. The ePUB Toolbox groups methods according to their use in *ex ante*, real time and *post hoc* evaluations. Methods employed in the *ex-ante evaluation* of programmes and policies include:

- Foresight studies
- Modelling and simulation
- Cost-efficiency techniques
- Cost-benefit techniques

Quantitative Methods employed in *Monitoring and ex-post evaluation* include:

- Statistical data analysis: Innovation Surveys and Benchmarking
- Modelling methodologies: Macroeconomic modelling and simulation approaches; Microeconometric modelling; Productivity analysis; and Control group approaches

Qualitative and semi-quantitative methodologies employed in *Monitoring and ex-post evaluation* include:

- Interviews and case studies
- Cost-benefit analysis
- Expert Panels/Peer Review
- Network Analysis
- Foresight/ Technology
6.6 What sorts of Measurement are we looking at in Innovation Programme Evaluation?

The ultimate rationale for an innovation programme is likely to be a matter of contributing to such objectives as increasing the competitiveness of a region, an industrial cluster, or a national economy; creating more (or better) jobs and employment opportunities; helping industry to move in a more environmentally sustainable direction, and so on. As discussed earlier, finding operational definitions for some of these terms, that can be translated into measurable indicators, may be no easy matter. Concepts like “competitiveness” and “sustainability” are contentious ones, for example.

When we have a fairly uncontentious concept, finding appropriate indicators may not be difficult. Though there may be debates about how we deal with part-time employment and temporary jobs, for example, employment is fairly readily assessed and employment levels are indeed captured in many routine statistical enquiries. Even in this case it may be helpful to provide data on full- and part-time jobs – or be clear when “full-time equivalent” indicators are being used. Where contentious concepts are being addressed, it is much more important to use multiple indicators. Each is only liable to address part of the issue at hand, and different indicators may be most relevant to different formulations of the concept. Multiple indicators also allow for more “triangulation” of the evidence, where we can examine how far trends are similar across different aspects of the issue at hand.

Additionally, (a) many of the developments that concern us in innovation programmes are inherently long-term ones – the impact may not be fully visible for years. Furthermore, (b) we are more concerned with lasting impacts – not with a short-lived improvement in performance.

The first consideration (a) means that it will often be necessary to work with indicators of “intermediate impacts”, that can reasonably be presumed to indicate that things are moving in the direction of long-term change. This will often apply, for example, to the economic performance of firms or industries. It is frequently the case that the benefits of introducing an innovation are slow to appear – the product takes time to gain a sizeable market, there is a learning period required for the new process to be implemented most effectively. (R&D can take even longer to have visible pay-offs.) Even after an innovation programme has been concluded, it may be necessary to rely on estimates of the likely effects of an innovation that has been introduced. (This may involve drawing inferences from earlier innovations, from statistical modelling, from examining the experience of vanguard innovators, from surveying programme participants as to their expectations – all of which forms of evidence and argument have their drawbacks.) The indicators in which we can have most confidence – the observed diffusion and uptake levels of the innovation, for example – are still fairly removed from the ultimate impacts with which we are concerned.

The same is true in respect of point (b). How much “behavioural additionality” (see Chapter 7 for more on this term) has really been achieved is a major
question. For example, has the programme only resulted in the adoption of a particular innovation, or has it led to a changed orientation among the participants? A single change in, say, CO₂ emissions, may be a valuable achievement, but are the organisations continuing to make improvements in their emissions year on year? Have they just adopted one innovation and assumed that this was sufficient for competitiveness, sustainability, or whatever other goal is aimed at? To determine such lasting, longer-term impacts, it will typically be necessary to examine changes in the capabilities of the organisations – whether they are devoting resources to exploring further innovations, for example.

Chapter 7 deals in more detail with questions of impact assessment and additionality.
6.7 What are the sorts of secondary data that may be accessed?

Some data may be available from existing sources: secondary data that can be used for evaluation purposes. For example, routine official statistics carry a great deal of material on economic activity – turnover, employment, productivity, and the like. These often provide baseline information against which we can assess changes, or a general statistical background for the study to use. Ideally, such material will be already assembled at the outset of an innovation programme, but the time-lags involved in preparing and publishing statistics means that often those available at any moment reflect circumstances several years earlier.

As well as data presented in the Programme Overview, we can expect data to be produced in the course of monitoring of the programme. Some of these should be data on projects, expenditures, and activities; some may cover immediate impacts of the programme. There may be quantitative information, or more archival documentation of activities (lists of names, for instance). Individual projects are also likely to generate data that can be useful for evaluation purposes, which again may be more or less discursive or quantitative. To the extent that this material has been organised and processed already, it can be secondary data for the evaluation study. If it has to be coded, classified, or heavily reworked, it may be best seen as primary data.

Data may be produced by external sources that can be useful for evaluation of programme impacts, though this is rarely sufficiently timely or attuned to programme issues to be of great relevance. It is not impossible that an industry or cluster study will be conducted that will tell us something of significance to the programme, though. In the UK in the mid-1990s, for example, an innovation survey carried out by the Confederation of British Industry enquired about the extent of knowledge of, and participation in, the Technology Foresight Programme that was then underway. This provided information that allowed us to assess how far the Programme was engaged with different sectors of industry. (The results suggested that it was finding less response from service sectors of the economy – see Miles, 1999) In Germany more recently, the Community Innovation Survey was used to determine how far there was an impact of government R&D funding on business R&D.

A problem that can confront users of secondary data is gaining access to material in a form that is useful for the evaluation. Statistics may be presented in too aggregated a form – we may need data disaggregated by industry, firm size, and so on. Data may be withheld on grounds of confidentiality – with statistics there may be rules that prohibit access to information that could conceivably allow individual responses to be identified, and with archival material there may be problems to do with naming of individuals, off-the-record comments, and the like. Consultants may be unwilling to share data that they have produced, for a variety of reasons. (E.g.: such data are regarded as financially valuable; suspect methods of producing data are being concealed; there are concerns that analysts may draw results other than those proffered by the consultant; etc.)
6.8 What are the Methods for generating data via primary research?

Secondary data will have been produced as primary data by some other party: in evaluation studies we are also usually in the business of producing some primary data, too. Different techniques typically produce different sorts of data, but there is not a one-to-one mapping. Questionnaire surveys are typically used to create large-scale quantitative data, but can also be carried out using quite small samples, and can be used to elicit open-ended, qualitative inputs. Interviews generally yield a great deal of qualitative data, but at least some of the material may take a quantitative form (and qualitative accounts may be coded into quantitative measures).

In the following discussion, when we discuss “new data”, we refer to material generated by observing phenomena in the material world – these phenomena may be behaviour we can see through participant observation of meetings, for example, or from the existence of records in filing cabinets or electronic storage. They may be statements (of opinion, belief, etc.) elicited from interviewees, survey respondents, workshop participants. This rather obvious point is being laboured in order to distinguish such data from the information that can be generated from, for example, simulation modelling of events. The status of information about people's actual behaviour or their stated opinions is rather different from that of information about what a computer model says. The latter is telling us about how the model processes data (that has been generated from empirical research or expert judgement). We consider this a form of data analysis and drawing of conclusions, rather than generation of new data for evaluation purposes.

Among the main tools for generating new data, then, are:

♦ Questionnaire Surveys
♦ Interviews
♦ Workshops and Focus Groups
♦ Observation
♦ Archival Analysis.

The illustration outlines an appraisal of the overall advantages and challenges of each technique. We provide a little more detail on each below. But two general points can be made immediately. First, it is practically always vital to engage in some pilot work before launching into full-scale data generation. There are many sad stories of survey studies ruined by errors in questionnaire design, or interview programmes thrown out by (typically) failure to address key issues adequately. Workshops need careful preparation, and even if a “dry run” cannot be achieved, some role playing may give a sense of what will and will not work effectively. It is wise to observe some events in an open-minded way before setting out to record and code specific activities. Similarly, archival analysis will typically require some experimentation with classifications and categories before we launch into a massive coding exercise.
Second, the issue of **sampling** is very often raised. Often we cannot hope to talk to everyone who was a participant or stakeholder in a project, or examine every document that was produced in the course of the work. The question then becomes one of knowing just how far the data we use are representative – representative of the whole population or of some subset with which we are particularly concerned. We will thus need a strategy to decide just where we will need to go for the data we want, and this strategy will have to be justifiable in terms of the evaluation questions that we are asking.
### Table 6.1 A Succinct Comparison of Data Generation Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall Purpose</th>
<th>Advantages</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaires, Surveys, Checklists</td>
<td>When need to quickly and easily get lots of information from people in a non threatening way</td>
<td>Can complete anonymously, inexpensive to administer, easy to compare and analyze, administer to many people, can get lots of data, many sample questionnaires already exist</td>
<td>Might not get careful feedback, wording can bias client's responses, are impersonal in surveys, may need sampling, expert doesn't get full story</td>
</tr>
<tr>
<td>Interviews</td>
<td>When want to fully understand someone's impressions or experiences, or learn more about their answers to questionnaires</td>
<td>Get full range and depth of information develops relationship with client, can be flexible with client</td>
<td>Can take a lot of time, can be hard to analyze and compare, can be costly, interviewer can bias client's responses</td>
</tr>
<tr>
<td>Documentation Review</td>
<td>When want an impression of how program operates without interrupting the program; is from a review of applications, finances, memos, minutes</td>
<td>Get comprehensive and historical information, doesn't interrupt program or client's routine in program, information already exists, few biases about information</td>
<td>Often takes a lot of time, info may be incomplete, need to be clear about what looking for, not a flexible way to get data, data restricted to what already exists</td>
</tr>
<tr>
<td>Observation</td>
<td>Gather accurate information about how a program actually operates, particularly about processes</td>
<td>View operations of a program as they are actually occurring can adapt to events as they occur</td>
<td>Can be difficult to interpret, seen behaviours can be complex to categorize, observations can influence behaviours of program participants, can be expensive</td>
</tr>
<tr>
<td>Focus groups</td>
<td>Explore a topic in depth through group discussion, e.g., about reactions to an experience or suggestion, understanding common complaints, etc., useful in evaluation and marketing</td>
<td>Quick and reliable, get common impressions, can be efficient way to get a wide range and depth of information in a short time, can convey key information about programs</td>
<td>Can be hard to analyze responses, need good facilitator for safety and closure, difficult to schedule 6-8 people together</td>
</tr>
<tr>
<td>Case studies</td>
<td>To fully understand or depict client's experiences in a program, and conduct comprehensive examination through cross comparison of cases</td>
<td>Fully depicts client's experience in program input, process and results, powerful means to portray program to outsiders</td>
<td>Usually quite time consuming to collect, organize and describe; represents depth of information, rather than breadth</td>
</tr>
</tbody>
</table>

Source: Carter McNamara (n.d.) Basic Business Research Methods
6.9 What is the Use of Interview-based methods for generating data in Evaluations?

Interviews have been described as “structured conversations”. They are face-to-face, or telephone-based, encounters, in which the evaluator asks a series of questions. Usually there is one informant, but there may be more than one person contributing in this way. (This can be especially useful where the interviewees have complementary information to present; where there are reasons to avoid presenting the same introductory orientation remarks over more times than absolutely necessary; where there are advantages in facilitating the interviewees’ learning from each other in the course of the interview. The corresponding disadvantage is that interviewees may be less willing to give full and open responses in the presence of other interviewees – this may be particularly the case where there are status differences, for example.) Usually there is one main interviewee – sometimes a partner takes notes, sometimes interviewers specialise in particular lines of questioning.

Interviews may be more or less structured: most common in evaluation research is the semi-structured interview. Here there is a list of set questions to be posed, but there is considerable flexibility to change the sequence of questions if (as often happens) the interviewee moves on in the course of replying to one question, to providing inputs bearing on a later question. It is more important to maintain the conversational flow than to follow a preordained order of topics. There is also flexibility in the shape of being able to introduce new lines of enquiry if interesting unanticipated issues arise – and to drop lines of enquiry that are yielding few returns.

Completely unstructured interviews are rare, but interviews where there is only a loose structure are most often used at the outset of a study, where the evaluator is seeking to gain a general picture of how the programme is structured and operating. This may be particularly relevant where the Programme Overview is absent or insufficient.

Highly structured interviews are essentially questionnaire surveys conducted verbally, with the interviewee relieved of the task of writing down responses. The advantage here is that often informants are unwilling to commit as much time and energy in writing comments than they are prepared to give in a face-to-face encounter. Most people find speech less taxing than writing (and, since it is less formal, they find it less categorical and definitive).

Interviews are most valuable when they are being used to generate detailed qualitative accounts. It is possible to include questions in an interview that ask people to provide quantitative estimates, to make judgements on Likert-scales, and the like. Large-scale telephone interviewing is often used in this way, with professional opinion poll or market research staff administering a questionnaire about whose contents and subject they know little. But the commitment of evaluator resources to interviews, especially face-to-face ones, means that such interviews are most effective when capturing the breadth and depth of qualitative material that surveys will largely miss.
The interview yields more than an array of data points that can only be related together statistically. It can elicit the interviewee’s account of the programme, a project, or at least the part of the programme or project with which he/she was involved. As well as describing processes and outcomes, the interviewee may offer us an understanding of why processes and outcomes were related together in particular ways; what explains anomalies and regularities; what the value of specific indicators might be; where to go for further evidence; and so on. People are naturally story-tellers, constructing narratives of their experiences. An evaluation report will be a narrative of sorts, itself. Interviews are an excellent technique for finding out what the narratives of participants and stakeholders are.

The big challenge that this poses for evaluators is less one of making sense of the data – the narratives are rich with meaning. The challenge is one of validating and integrating these narratives, and the detailed qualitative data which is articulated within them. This will be discussed further in the questions on data analysis to be dealt with later in this chapter.
6.10 How and when should Focus Groups and Workshops be used for generating data?

The larger and longer-lasting evaluation studies will often have a Steering Committee that helps in the specification of the evaluation study, comments on design and interim results, makes suggestions as to reporting, and so on. The Steering Committee is likely to be composed of stakeholder and especially end-user representatives, programme managers, people from the evaluation community, and so on. However, it is possible and sometimes well worthwhile to create workgroups drawing mainly or exclusively on participants and perhaps some other highly involved stakeholders. These groups will usually only meet on one occasion (though in principle they could meet more often).

They may be used early in the course of an evaluation, where they may prove useful for raising issues about the programme that the evaluation will need to deal with, where participants can comment upon (for example) proposed impact indicators, and so on. The main problem with setting up such groups early in the course of an evaluation is that the evaluators may have little information of their own to share with the programme participants, it may be difficult to define exactly how the workshop should operate, and there may be correspondingly less to attract involvement from the desired informants. Later in the course of an evaluation, a workshop can be an ideal setting in which to present and test out preliminary results and conclusions. It also can provide opportunities to explore apparent disagreements and divergent accounts. There are circumstances where informants will proceed to much greater depth in the course of dialogue with each other than they are likely to do in an interview setting. Workshops can generate interchanges between participants that press them to justify and explicate their views in detail. Workshop approaches, finally, are often valuable where more participatory approaches to evaluation are being attempted, and they are a staple element in foresight studies (see below).

There are numerous variants of focus group and workshop, drawing on a range of specific techniques (including brainstorming, role-playing, simulation gaming, scenario analysis, and various creativity techniques). While the traditional tools were flip-charts and whiteboards, we are seeing increasing use of computer support within face-to-face meetings. (The latter have the advantage of “capturing” a great deal of information electronically, and allowing several participants to input views simultaneously.) In either case, the critical issue is to plan the workshop process well in advance, and to make sure that facilitator(s) are able to keep the process moving along, return discussion to the subject at hand if it veers off-topic, resolve conflicts as they arise, and maintain the interest and involvement of participants. Material generated in the workshop should be recorded and transcribed as thoroughly as possible – though this does mean that rather a lot of qualitative data may require processing subsequently.
<table>
<thead>
<tr>
<th>Factors to consider</th>
<th>Use focus groups when...</th>
<th>Use indepth interview when...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group interaction</strong></td>
<td>interaction of respondents may stimulate a richer response or new and valuable thought.</td>
<td>group interaction is likely to be limited or nonproductive.</td>
</tr>
<tr>
<td><strong>Group/peer pressure</strong></td>
<td>group/peer pressure will be valuable in challenging the thinking of respondents and illuminating conflicting opinions.</td>
<td>group/peer pressure would inhibit responses and cloud the meaning of results. Color</td>
</tr>
<tr>
<td><strong>Sensitivity of subject matter</strong></td>
<td>subject matter is not so sensitive that respondents will temper responses or withhold information.</td>
<td>subject matter is so sensitive that respondents would be unwilling to talk openly in a group.</td>
</tr>
<tr>
<td><strong>Depth of individual responses</strong></td>
<td>the topic is such that most respondents can say all that is relevant or all that they know in less than 10 minutes.</td>
<td>the topic is such that a greater depth of response per individual is desirable, as with complex subject matter and very knowledgeable respondents.</td>
</tr>
<tr>
<td><strong>Data collector fatigue</strong></td>
<td>it is desirable to have one individual conduct the data collection; a few groups will not create fatigue or boredom for one person.</td>
<td>it is possible to use numerous individuals on the project; one interviewer would become fatigued or bored conducting all interviews.</td>
</tr>
<tr>
<td><strong>Extent of issues to be covered</strong></td>
<td>the volume of issues to cover is not extensive.</td>
<td>a greater volume of issues must be covered.</td>
</tr>
<tr>
<td><strong>Continuity of information</strong></td>
<td>a single subject area is being examined in depth and strings of behaviors are less relevant.</td>
<td>it is necessary to understand how attitudes and behaviors link together on an individual basis.</td>
</tr>
<tr>
<td><strong>Experimentation with interview guide</strong></td>
<td>enough is known to establish a meaningful topic guide.</td>
<td>it may be necessary to develop the interview guide by altering it after each of the initial interviews.</td>
</tr>
<tr>
<td><strong>Observation by stakeholders</strong></td>
<td>it is desirable for stakeholders to hear what participants have to say.</td>
<td>Stakeholders do not need to hear firsthand the opinions of participants.</td>
</tr>
<tr>
<td><strong>Logistics</strong></td>
<td>geographically an acceptable number of target respondents can be assembled in one location.</td>
<td>respondents are dispersed or not easily assembled for other reasons.</td>
</tr>
<tr>
<td><strong>Cost and training</strong></td>
<td>quick turnaround is critical, and funds are limited.</td>
<td>quick turnaround is not critical, and budget will permit higher cost.</td>
</tr>
<tr>
<td><strong>Availability of qualified staff</strong></td>
<td>focus group facilitators need to be able to control and manage groups</td>
<td>interviewers need to be supportive and skilled listeners.</td>
</tr>
</tbody>
</table>

Note: "indepth interviews" in the original text refers to un- or semi-structured interviews.
6.11 How and when should Observational Methods be used for generating data?

Observations are often employed in the course of real-time evaluations, and sometimes in interim evaluations and Ex Post evaluations where there is an opportunity to be involved in activities that are following on from the programme. The typical ways in which evaluators are liable to be engaged in evaluations involve participation in programme activities, such as meetings, conferences, dissemination events, or similar activities where programmes may have impinged, such as trade fairs or industry events. (In this way the observation of innovation programmes is likely to differ from that of, say, educational programmes – in the latter case evaluators may find themselves sitting in classrooms or teachers’ common rooms. There are greater problems of observers influencing the thing being evaluated in such cases; but correspondingly they will be examining more directly the work that is being influenced by the programme. It is not unknown for evaluators to be taken on a tour of factories, offices, or workshops, or to be given a demonstration of software or a chance to listen in to helpline telephone calls, however. Usually, unless there are ethical problems involved, such opportunities are well worth seizing – they can provide rich insights into what the programme has being attempting to achieve, and what has been happening in the course of events.)

Observational research provides the evaluator with the opportunity to see how the programme is operating and/or being responded to, in real-life circumstances. Typically, these settings will be fairly naturalistic, with limited influence on participants’ behaviour from the presence of the evaluators.

In smaller meetings, it is possible to observe a fairly high proportion of what is happening. The challenge may be one of recording this, when many people are present and interacting. In larger meetings, it will be necessary to be highly selective about what to observe and record, and decisions will need to be made about how to allocate time across different parts of the environment. (For instance, an evaluator who is observing attendance at a particular set of stalls in a trade fair will not be able to be simultaneously interviewing passers by, or visiting other stalls to see if they show any influence from the programme.) It is important to strive to validate the impressionistic conclusions that would be drawn from more casual observation – for instance, a count of the number of people attending a meeting or stall is far more useful than the view that “not many” people were in evidence. The sorts of thing one is seeking to observe, and the sorts of data that will be required, require attention at an early stage.

Observational approaches may be extremely valuable as a source of background information for evaluators, even if the information generated does not form data that can be directly drawn on and cited. Taking part in events may give a good impression of what activities are underway and how they are being conducted and received, and also, perhaps, how they fit into the wider innovation scene. Systematic information may be developed from observation of events. Informal observation of one event may provide insight as to key issues and possible indicators to elaborate in future events.
6.12 How and when should Surveys be used for generating data?

Questionnaire Surveys allow for a large sample of informants to be reached in a relatively cost-effective manner. While qualitative interviews typically provide in-depth information about a few cases, quantitative surveys typically provide more basic structured information about a relatively large number of cases. (As with most statements about methods, there are many exceptions to these rules.)

Surveys for evaluation studies were until recently mainly conducted by postal means: intended informants were sent a copy of a questionnaire by mail and invited or instructed to complete it. This method runs into the problem of low response rates, unless there are strong inducements for providing answers; one way of increasing response rates has been to follow up the initial questionnaire with frequent reminders, including telephone calls. Telephone-based interviewing has become a profession in its own right, and skilled staff using supportive computer systems are often employed in interviewing industrial and policymaking informants. (These capabilities are now often provided by specialised survey research firms.) Surveys are also often conducted using email questionnaires and online questionnaires. These have the advantage of being easy to distribute to large numbers of target informants, and providing material in electronic form that is immediately ready to process. However, the problems of low response rate apply to these as to postal questionnaires.

Usually, a questionnaire in an evaluation study will address topics such as: the characteristics of the informant and the organisation they represent; the organisation's involvement in the programme in question (and in any other relevant programmes); what the performance of the organisation has been and whether there are believed to be impacts of various kinds associated with the programme (especially those that suggest behavioural additionality); whether there are insights as to wider impacts of the programme (e.g. on networks which they belong to).

The design of questionnaires has been discussed in depth in many methodological texts, and this is not the place to discuss, for example, different types of scale, or even the general rules for formulating questions that such texts spell out. Designing a questionnaire forces us to confront the question of what indicators we are employing in a very stark way. In order to keep the questionnaire to a reasonable length, difficult decisions may need to be made as to which of several lines of enquiry need to be pursued.

The other major issue that confronts us in survey research is that of sampling. It can be possible to send a survey or address a telephone interview to every participant in a small programme, but with larger programmes the numbers of participants and stakeholders may be unmanageably large. Random sampling is not always the best approach to take, since we may need to stratify the sample so as to be sure of capturing the experiences of some key types of participant who are represented in small numbers. Sampling strategies are exhaustively discussed in many of the guides and texts we have mentioned above.
**Figure 6.1 Key Issues in Sampling**

<table>
<thead>
<tr>
<th>Key Term</th>
<th>Definition</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>“Universe”, Theoretical Population</strong></td>
<td>The entire set of cases (these are liable to be organisations or individuals) to whom the results are to be generalised.</td>
<td>In much social research the research will be attempting to make a generalisation as to what the value of a parameter or strength of a relationship might be. In evaluation research, we are more often asking what the effect of an intervention was on specific cases – but we are often interested in whether such effects might be generalised.</td>
</tr>
<tr>
<td><strong>Study Population</strong></td>
<td>The set of cases to which the researchers in principle could have access.</td>
<td>A list of programme participants will usually be available, though some individuals, organisations and addresses may no longer operate. The range of people likely to be indirectly influenced by a programme may be much harder to establish, and it may be necessary to go to a directory of businesses or other organisations.</td>
</tr>
<tr>
<td><strong>Sampling Frame</strong></td>
<td>The set of cases which the researchers attempt to access.</td>
<td>This frame is based on decisions the researchers make about how many members of the study population to attempt to contact, and on what basis (e.g. random or stratified sample) these will be selected.</td>
</tr>
<tr>
<td><strong>Sample</strong></td>
<td>The cases on which the researcher is actually able to generate useful data.</td>
<td>Data may be more or less complete for different cases. Decisions are needed as to how complete a response is required for the case to be regarded as an effective member of the sample, rather than a non-response?</td>
</tr>
</tbody>
</table>

6.13 What are the Methods for structuring and exploring interventions in the course of generating data?

In natural sciences and some types of social research (such as psychology), the “gold standard” in data production is controlled experiments. For an innovation programme, this would mean acting in the design of the programme itself, not just in the evaluation. The design should randomly decide who should and who should not participate in the programme, and then evaluators could compare the experiences of the two groups in order to assess the impacts of the programme on them. Oh, and there should be no communication between the two groups, and efforts would need to be made to ensure that there were no indirect influences taking place through networks or supply chains…. It is clear that the ideal(ised) controlled experiments will not be commonplace in innovation programme design and evaluation.

Data production and analysis in evaluation studies can take opportunities to make some sorts of comparisons, however. Surveys, for example, can deliberately build in comparison groups. **Comparison groups** are not true control groups, where an experimenter has selected who and who should not fall into the group receiving the treatment. They are groups that have not been directly involved in the programme we are evaluating, and that are as close as possible to those that were involved in the programme. There may be a process of forming “matched pairs” that correspond to the different participants sampled for attention, or creation on a comparison group on the basis of statistical treatment of a large database of organisations on whom relevant data are available. In the US’ Manufacturing Extension Services Program such comparison group surveys were used in evaluations of (inter alia) the importance of the Program compared to other research funding programs, and some of the key impacts of the Program. Shapira and Kuhlmann (2003) report that when comparison group surveys were used, the results were more mixed than when surveys of participants alone were used.

Statistical methods – multivariate analyses - allow for the processing of data on samples of bodies differing in many characteristics, “controlling” for the effects of parameters (as far as his is possible given sample size and correlation between parameters). A more ad hoc method often employed by researchers is to set up “matched pairs” to compare, cases that are as alike as possible except in terms of the critical variable of interest (and its correlates or consequences – such as the effects of a programme intervention).

In the course of action research and participatory approaches, the interventions made are rather different ones, reflecting the researchers’ engagement with the objects of study.
Box 6.3 The Manufacturing Extension Services Program

The Manufacturing Extension Partnership (MEP) provides manufacturing services to small and medium-sized manufacturers, in so doing, aims to upgrade the performance and competitiveness of these firms. The kinds of services offered include information dissemination, assessment, problem-solving, referral, and training. The MEP is a collaborative initiative between the US federal and state governments that also involves non-profit organisations, academic institutions, and industry groups. The MEP’s federal sponsor is the National Institute of Standards and Technology (NIST), within the US Department of Commerce. While the federal sponsor, NIST, performs an ongoing funding and co-ordinating role, the states are also major funders of manufacturing extension services. One of the interesting features of the US manufacturing extension system is the considerable effort placed on performance measurement and evaluation.

Findings from MEP evaluation studies
About 30 evaluation studies of the program have been produced by a variety of authors over the last five years or so. A variety of methods have been used in these studies, with some using more than one method. While the most common method was to survey customers, it is worth noting that the next most common method was a survey that also used a comparison group. Case studies were the third most frequently used approach. When control group surveys are implemented, the results that emerge from the portfolio of such studies show more mixed results than surveys without comparison groups.

Evaluation insights from MEP
The MEP case illustrates that evaluation involves far more than ‘collecting information’. Resources need to be allocated to analyse the data in ways that are useful to program managers, staff, sponsors, and outside stakeholders. Results (whether good or bad) need to be explored to understand what caused what and why and how much was attributable to the program. Good case studies are often worthwhile here. The MEP case emphasises that evaluation should be tasked not only with justifying the program to funding sponsors, but also with promoting learning and program improvement. If this is to succeed, it is also important that an evaluation community be stimulated and supported (through training, workshops, and electronic interaction) to allow the development of evaluation capabilities, the application of robust and innovative methods, and the active dissemination of findings. This community should have opportunities and requirements to engage in dialogue about findings, needs, and implications with program managers, sponsors, industry representatives, and policymakers.

6.14 What are the Methods for analysing data?

Data analysis needs to be planned at an early stage, because ideas about what sorts of analyses need to be undertaken are extremely important in telling the researcher what sorts of data to generate and how to code and process it. Many people have set out to amass huge volumes of data with little thought of how precisely it can be used, and then found themselves having to climb a steep learning curve in order to extract sense from it. There are also often many regrets about how the questionnaire was designed, the data were coded, how the material was or was not entered into a computer system. Thinking about data analysis at an early stage of the work can help (a) make sure that you are asking the right questions, with the right instruments and (b) make sure that you are capturing the information in a way that you can use it efficiently.

**Quantitative Data Analysis** is a matter of statistical methods and means for visualising data. There are many useful guides here, and a number of well-known computer programs for generating results using routines ranging from the very basic to the highly sophisticated. What these programs do not do well is to warn the user about basic errors that can be made – such as the use of significance tests in inappropriate circumstances (very common in applied social research).

**Qualitative Data Analysis** is another matter – until fairly recently there were numerous guides to how to conduct interviews, but very little about how to use the masses of text that can be generated from them. Knowing whether the conclusions drawn from a set of interviews are robust ones was very much a matter of faith in the researcher. More recently there have become available software packages for coding and analysing interview and other text-based data, and several extremely useful texts reviewing the range of methods and good practices here.

Techniques such as “meta-analysis” are becoming more widely used as ways of drawing conclusions from a number of individual studies. (For example, Kingsley and Klein (1998) used a meta-analysis methodology to analyse 123 case studies of industrial networks, concluding that network membership can be built with the sponsorship of parent organisations and with public funding - but the networks that generate new business are associated with private sector leadership and at least some private sector funding.)

A good place to begin for data analysis tools is to examine the various handbooks that are available – many published by Sage publications, that has been very systematic in the area. A useful website on evaluation with many links to data analysis and other methodological issues is at: http://www.edc.org/hec/eval/links.html
Sources for quantitative data analysis:

Useful Canadian guide to presenting and analysing statistics: http://www.statcan.ca/english/edu/power/toc/contents.htm

An online version of an introductory guide (from the US GAO, that also offers other evaluation-relevant resources):

Sources for qualitative data analysis:

An online journal is at: http://www.qualitative-research.net/

Sources for software: http://www.uwm.edu/People/brodg/CQDAbib.html

The classic guide to the topic is: Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.) London: Sage

Patton, M. Q. (1990) *Qualitative Evaluation and Research Methods* (2nd ed.) Beverly Hills, Sage,
6.15 What are the advantages of combining qualitative and quantitative approaches?

Quantitative data are powerful means of expressing the extent and scale of issues and impacts, and are more readily converted into monetary values as required in cost-benefit analysis and efforts to estimate rates of return on investments. Qualitative data can provide insight into processes as they have been experienced. Some policymakers require quantified results, preferring figures and graphs; others prefer narratives and case studies. One of our interviewees, based in Switzerland, pointed out that:

“There is a certain tension between evaluators preferring a qualitative approach and those who prefer a quantitative one. This tension leads to an uncertainty for policy-makers as to the appropriate evaluation strategy they should pursue. A more intensive discussion among the different “schools” of evaluation thinking would help all involved to realise that the methods could be used complementarily to each other”.

In some cases, there are strong institutional pressures in favour of one or other approach. Thus, the need in the USA, the role of evaluation is very much geared to justification of government spending and to use evaluation as providing evidence for the impact of programmes. In the US this has led to emphasis on a more quantitative approach of measurement of performance in terms of targets: the US Government Performance Act described below has imposed this onto agencies in each policy area, for instance. (See Box 6.3)

Purely quantitative analysis is not appropriate for some types of innovation programme, however. For example, promotion of cluster formation is an important policy objective in many countries. In order to assess any change in the volume and quality of linkages between the stakeholders in a cluster, evaluators may need to examine interactions with a mixture of approaches (including surveys, interviews, and some attempts at network mapping).

Frechtling and Sharpe (1997) focused on “mixed methods”, referring to a combination of qualitative and quantitative methods in evaluation studies. They review arguments for and against each approach, and consider the arguments for combining the two as being:

♦ Improved triangulation of results, in that very different sorts of technique are being used to gain a view of the same phenomena (or different aspects of the same phenomena), thus providing better insights into the complex situation.

13 The handbook concerns evaluation of teaching and education programmes, but much of the discussion is more generic.
Improving measurement and indicators (for example, by using focus groups to identify possible questions for a survey, or to comment on drafts of such questions).

Identifying points where the evaluation framework may need to be extended – topics that need to be covered, inconsistencies in results that need to be addressed, and the like.

To these we would add:

Ability to combine in-depth insight and process understanding with a sense of the broad landscape and the extent to which individual cases can be generalised.

Ability to generate different types of account in such a way that different audiences can be addressed.
Box 6.4 The Government Performance and Results Act (GPRA)

In 1993, the U.S. federal government passed the Government Performance and Results Act (GPRA). This Act was modelled on similar legislation already in place in the states, and in several other countries. The Act requires every agency to make a strategic plan every three years; translate it into specific annual performance plans with quantified targets on performance indicators; and report every year on whether the targets were met.

Because the industry-oriented RTD programs were already heavily evaluated, most federal agencies have been able to use existing performance indicators to respond to the GPRA requirements. NIST, for example, has drawn on the ATP case studies in its performance report. The need for performance indicators at the federal level has also shaped what happens in the states. NIST's Manufacturing Extension Program has shaped the list of tracking data and performance indicators used in many state extension programs. Following the pattern in the U.S. system generally, however, GPRA has been implemented highly unevenly. Many of the programs listed above are too small to appear on their own in GPRA strategic plans or performance reports. Even the federal programs in this category go through different sections of the President's Office of Management and Budget, and through different Congressional committees. Each of these offices and committees has tended to develop its own expectations, which vary significantly. Elements of industry-oriented RTD schemes that do not take the form of programs with budgets (e.g., tax credits) are not subject to GPRA examination at all. In terms of impact, the evidence to date does not indicate that GPRA performance measurement is having a significant impact on resource allocation by legislatures. At state level, performance indicators seem to be reported rather perfunctorily, and anecdotes of success have more effect. At federal level, a new report on GPRA implementation identifies lack of use of the indicators as a de-motivating factor for many agencies. Lack of connection to resource allocation does not mean that the Act as a whole is not having an impact. Many observers note that strategic planning and performance assessment is forcing agencies to think through their missions more carefully and make sure what they are doing on a day-to-day basis supports those missions. Thus shifts in internal resource allocation may be a better indicator of the impact of the law than Congressional changes. Likewise, development of indicator systems, rather than the actual levels of the indicators themselves, may have the strongest impact on agency operations.

Source: An international review of methods to measure relative effectiveness of technology policy instruments (July 2001), Report by Technopolis (Netherlands), Georgia Institute of Technology (USA) and University of Sydney (Australia).
Box 6.5 Mixed-Method Evaluation: the Case of EUREKA

The first evaluation of the EUREKA initiative (Ormala et al., 1993) provides a helpful example of combining both qualitative and quantitative methodological approaches. Created as an intergovernmental Initiative in 1985, EUREKA aims to enhance European competitiveness through its support to businesses, research centres and universities who carry out pan-European projects to develop innovative products, processes and services. EUREKA offers project partners rapid access to a wealth of knowledge, skills and expertise across Europe. Through a EUREKA project, partners develop new technologies for which they agree the Intellectual Property Rights and build partnerships to penetrate new markets.

The evaluation analysis was based on three levels of information utilising different both qualitative and quantitative methodological approaches.

Level 1 was the compilation of data on the basic characteristics of projects and participants, carried out through a compilation of data held by national administrations and other documentary sources.

Level 2 analysis involved the identification of direct and indirect economic effects of projects. This involved sending separate questionnaires to industrial and non-industrial participants in the sample frame. The response rates seem very good, averaging around 57%.

Level 3 information was collected through an in-depth analysis of a representative sample of projects. A representative sample of 73 projects (17.5% of total) was selected to be interviewed in depth of which 70 were eventually completed. These issues could only be addressed through the combination of both qualitative and quantitative approaches, as outlined above.
6.16 What is Cost-Benefit Analysis?

Cost-benefit analysis (CBA) is based on the simple idea that a project or programme can be assessed in terms of the classes of benefits and costs that are associated with it. These benefits and costs are put into a common accounting framework by assigning monetary values to them. Then the relative merits of different projects or programmes, or the overall efficiency of a single option, can be assessed arithmetically. This simple idea requires a number of rather complex – and often contentious – steps in order to put it into practice:

♦ (Ex post) measurement or (ex ante) estimation of the costs and benefits – this encounters the familiar problem of limitations in being able to account for the complete range of costs and benefits (including unexpected outcomes).

♦ Assigning monetary values to these – particular controversy has arisen around (a) valuation of human health and life (attaching a money value to an individual life can cause deep unease) and (b) valuation of environmental sites and historical artifacts (where there may be irretrievably change or destruction of assets whose value to future generations is unpredictable).

♦ Deciding how to deal with uncertainties – cost-benefit analyses have often been conducted on the assumption that a “most likely” future is the only future we can or should plan for.

♦ Deciding how to deal with issues of equity and uneven distribution of costs and benefits that are liable to arise – there have been controversies as to the unequal treatment of different people, for example the assumption in cost-benefit analyses of transport programmes that the time of executives should be assigned a higher value than that of housewives. Generally, CBA does not attempt to assess impacts on different groups, assuming that market valuations will provide sufficient information to encompass overall societal impact. There could, however, be some assessment of differential impacts built in, though other approaches are more often applied to such issues.

♦ Determining how to discount future costs and benefits – while the standard economic discount rates can be applied, these can lead to very small importance being attached to long-term costs or benefits.

Cost-benefit approaches fit decision-making needs and routines so well that they are deeply engrained in institutional practice. As long as they are not seen as the complete answer (rather they are one of a number of approaches to evaluation), and as long as there is adequate acknowledgement of uncertainties and debatable assumptions (especially as regards monetisation), they may prove helpful in evaluation. The main issue is whether we can and indeed should attempt to reduce policy objectives and decisions to monetary parameters.
### Table 6.3 Types of Private and Social Costs and Benefits

<table>
<thead>
<tr>
<th></th>
<th>Individual Partners</th>
<th>Programme Sponsor (Government)</th>
<th>Society</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benefits</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>In Net Earning Profit</td>
<td>Tax Revenues</td>
<td>Increase in National Income (direct effect)</td>
</tr>
<tr>
<td></td>
<td>Additional Benefits from Transfers</td>
<td>Decrease in other Subsidies</td>
<td>Spill-over Effects (indirect effect)</td>
</tr>
<tr>
<td></td>
<td>Non-Economic Benefits</td>
<td>Tax Costs</td>
<td>Opportunity Costs (Cross Earnings from Other Potential Programmes)</td>
</tr>
<tr>
<td></td>
<td>Opportunity Cost of Participation</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Costs</strong></td>
<td></td>
<td>Project &amp; Administration Costs</td>
<td>Programme Costs</td>
</tr>
<tr>
<td></td>
<td>Direct Participation Cost</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Loss of Subsidies from Other Programmes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Table 12, in Wolfgang Polt and Birgit Woitech “Cost Benefit Analysis”, section 4.7 of the ePUB report (European Commission, 2002). This source is recommended for a detailed and cautious appraisal of CBA as applied to research programmes, with a helpful bibliography of approaches to CBA.
6.17 How can Cost-Benefit Analysis be Used?

Polt and Woitech (2002, p.144) usefully summarise the steps in Cost-Benefit Analysis as follows:

**Step 1: Identification and separations of relevant costs and benefits**

[...] costs and benefits can either be direct or indirect, tangible or intangible, internal or external. Problems can arise from costs or benefits that are not assigned to a certain project or program.

**Step 2: Monetary valuation of costs and benefits**

All relevant costs and benefits should be quantified in monetary terms. Where sufficiently perfect markets exist market prices can be used to determine the impacts. Difficulties arise whenever the market prices do not reflect the true marginal cost or benefits. This is often the case for intangibles and external effect. As described in section I shadow prices should be calculated to value these effects.

**Step 3: Discounting**

Since costs and benefits occurring at different times they should be discounted at an appropriate rate to make them comparable. Choosing the right discount rate is an important task. Therefore sensitivity analyses are recommended in order to systematically vary uncertain parameters to see what effect they have on the outcomes.

**Step 4: Aggregating**

After discounting the various effects of a project or program net present values, benefit-cost ratios or internal rate of return can be computed.

In terms of contribution to evaluation more broadly, *ex ante* CBA should be a useful tool for development of successful policies and programmes. This is more important than the accuracy with which it forecasts *ex post* costs and benefits. Likewise, ex post CBA is not mainly a matter of assessing the accuracy of ex ante estimates, but a source of insight for future policy. Undertaking a CBA is not just a routine matter of creating a spreadsheet to inform us about the “bottom line”; it can help organisations understand the key sources of uncertainty, what risks there are, and so on.

NERA (2004), reviewing lessons learned about the approach in a highly risk-prone policy setting\(^{14}\), concluded that: “where significant uncertainty exists, scenario analysis should be used to test the policy under a range of possible outcomes; and in publishing the CBA, [the sponsor] should identify any significant areas of uncertainty or risks to the success of the policy, including some illustration of the range and likelihoods reported…” (p21). Ex post CBA should be undertaken where: “there was significant uncertainty about the impact of a policy before Implementation; similar interventions are proposed in other areas; and ongoing costs are high relative to sunk costs”. NERA recommended “that ex post CBA focus on investigating specific policy impacts rather than necessarily attempting an evaluation of the total net impact (i.e. attempting to reproduce an ex ante CBA on an ex post basis)” (p24-25). Though these conclusions come from a different policy arena, they seem broadly applicable to innovation programme evaluation.

\(^{14}\) The study mainly concerned regulatory impact assessment concerning food standards, though attention was paid to examples of other circumstances.
Box 6.6 Social cost-benefit analysis for evaluating R&D programmes: an Australian example

“The basic methodology is one of a social cost-benefit framework, building on the important (but neglected) work of Fölster (1991). Much of the usefulness of the framework stems from the way in which it forces the evaluator to be specific about where and how any benefits or costs are likely to be manifested, as well as enumerating, albeit fairly crudely, the likely magnitudes.

The cost-benefit framework gives an appearance of precision and a scientific basis which is only incompletely founded. Many of the essential parameters for actually enumerating the benefits of innovation subsidies are not available with great accuracy, if at all. This implies that some degree of sensitivity analysis should be part of any evaluation, and, in the section on “Cost/benefit results”, we explore how the results vary as we alter the parameters.”

\[
\text{NSB} = \text{MARGIN} + \text{SPILLOVER} - \text{LEAK} - \text{MEB} - \text{ADMIN} - \text{COMPLIANCE} - \text{RENTSEEK}, \text{ where}
\]

\[
\text{MARGIN} = (\text{PB}^{\text{subsidy}} - \text{PB}^{\text{market}}) = i \times \text{ER} (r^{\text{subsidy}} - r^{\text{market}})
\]

is the difference (if any) between the private post subsidy rate of return \((r^{\text{subsidy}})\) on induced R&D \((i \times \text{ER})\) compared with alternative uses of those resources \((r^{\text{market}})\). \(\text{ER}\) is eligible R&D, and \(i\) is the inducement rate (equal to 0 if there is no inducement and to 1 if all of the R&D undertaken by a firm is new);

\[
\text{SPILLOVER} = \theta \times i \times \text{ER}
\]

are spillover benefits from induced R&D (at a rate of \(\theta\) per dollar of induced R&D). Spillover returns are those benefits derived from R&D spending that are not captured by the private investor;

\[
\text{MEB} = m \times s \times \text{ER} = m \times \text{R}
\]

is the marginal excess burden of taxation, i.e. the social costs associated with raising taxes in order to finance reductions in revenue due to business expenditure on eligible R&D. MEB is equal to \(m \times \text{R}\) where \(m\) is the marginal excess burden per dollar of tax revenue forgone and \(R\) is the tax revenue forgone.

\[
\text{LEAK} = \text{LEAK}
\]

is leakage of tax benefits to overseas shareholders. Such leakages only occur on the transfer element of the R&D. ADMIN are the government administrative costs of the program including the purely administrative costs of raising the tax revenue (which are not included in the MEB). COMPLIANCE measures the business compliance costs associated with the program. Finally, RENTSEEK represents the resources expended by firms and their representatives in lobbying for more generous provisions or in trying to resist the erosion of benefits.”

Source: Quoted from Lattimore 1998
6.18 Do the Methods we use vary according to the type of Evaluation?

Methodological principles remain much the same whatever sort of evaluation we are adopting, though some specific issues arise when, for instance, we are adopting more participatory approaches. (For example, there may be some trade-off between rigour and participation, though this apparent conflict is probably somewhat overstated.) Attention also needs to be addressed to whether we are engaged in ex ante, real-time, or ex post evaluation, since the different circumstances involved at these different time points are liable to affect what sorts of data we can employ and what methods are therefore available.

We shall consider these three types of evaluation in turn below, but can already compare and contrast them in broad terms:

♦ Ex Ante evaluation can have no actual impact data to work with, unless it can build on the results of pilot studies.
♦ Real-Time evaluation can utilise observations and examine processes as they are unfolding.
♦ Interim evaluation may be able to do some of these things that real-time evaluation can,
♦ Ex Post evaluation can examine impacts, though these usually are likely to be short-term and immediate ones.

Many of the same methods can be applied in the different types of evaluation, to the extent that similar sorts of data are available.
6.19 What Methods are Useful in Ex Ante Evaluation?

Typically, _ex ante_ evaluations are used to provide answers to questions of uncertainty and risk – what are the likely effects and impacts of this policy intervention? What is there for this programme investment - is it worthwhile in terms of value for money, in terms of achieving the desired (social, environmental, economic) results? At the time of an ex ante evaluation, we do not have any data on impacts – which is one reason for the authors of this Guide only reluctantly using the term “evaluation” for these activities – unless there have been pilot trials to draw on, or the programme is very closely based upon an earlier programme. Even if there are data such as from pilot trials or earlier programmes, it is unlikely that the results can simply be generalised to the current case. (Pilots, in particular, often give rise to high expectations that are subsequently depressed: it seems that initial energy and enthusiasm yields dividends, but that this is hard to maintain over the course of a larger programme. Also, often the recruits to pilots – and this may also apply to precursor programmes – are the “low hanging fruit”, the cases where success is most likely.)

Given the lack of impact data for the programme itself, ex ante evaluations will typically involve some more or less sophisticated modelling approach. In practice, this has sometimes been a matter of the application of cost-benefit analysis and related techniques, such as attempting to estimate the direct and indirect impacts of a programme through input-output tables. More often, rather less formal (and restrictive) methods are often used. The innovation process is known to be a complex one, with different innovations following patterns that vary considerably in detail, even if we can frequently employ S-curves to represent diffusion rates, and so on. While much policy analysis implicitly relies on a linear model of innovation, that has been widely discounted by researchers in the field, more complex analyses of innovation systems and industrial clusters are informing programme development – and ex ante analysis – in the institutions with a more advanced culture of evaluation. Often the complexity of the relationships being examined – and intervened in – is such that expert judgement methods are employed, such as the use of Panels and working groups in Foresight studies.
Box 6.7 Methods used in Ex Ante Evaluation

The *ex ante* evaluation of programmes and policies can employ a number of methodologies, which include:

- **Foresight studies**: These employ a structured consensus-building methodology based on the judgements of experts. In short, they provide an insight into most evident social, economical and technological development opportunities and outcomes available to policy planners.

- **Modelling and simulation**: These use quantitative modelling approaches to derive a single or a series of potential scenarios (through the modification of key variables) in order to estimate the socio-economic impacts of policy interventions.

- **Cost-efficiency techniques**: Again, these use quantitative, or semi-quantitative approaches to assess the costs and benefits associated with the policy intervention.

- **Cost-benefit techniques**: As indicated by its name, this approach is used to compare, in monetary terms, the range of social and private costs and benefits of a programme in order to ascertain that the balance between costs and benefits is a positive one. The technique can also be adapted to incorporate uncertainty and risk.

Source: Polt and Rojo, 2002 – RTD Evaluation Toolbox
6.20 What Methods are Useful in Monitoring?

In this Guide, we take ‘monitoring’ to be the process of examining programme activities during the course of the programme. It is more of a routine management activity than a real-time evaluation. Essentially, it involves examining activities and expenditures to establish whether these are proceeding according to plan, whether problems are being encountered, and so on. In some cases, monitoring may take the broad objectives and activities of the programme for granted, and will not be concerned with probing the effects and impacts of the programme except insofar as these are affecting planned progress. However, monitoring offers a valuable opportunity to capture, routinely and close to the moment that it is generated, information of relevance to the outputs, objectives, performance and progress of the programme. By doing so it avoids or, at least, facilitates the often resource-intensive process of data gathering necessitated by many ex post evaluations.

As such, monitoring may employ a range of qualitative, quantitative and econometric approaches, similar to those employed in ex post evaluation, to achieve its primary objectives of information collection. For example, apart from the routine collection of information of routine output and activity data (such as publications, projects undertaken, utilisation of resources, formation of network connections, etc.), more analytical tasks may be conducted, such as interviews or surveys of programme participants to review unidentified process outcomes, surveys of programme beneficiaries to check on unanticipated benefits, etc.
Box 6.8 Methods for Monitoring

The AmeriCorps Toolkit (2003), provided the following assessment of advantages and constraints of some of the most prevalent monitoring methods:

<table>
<thead>
<tr>
<th>Method</th>
<th>Characteristics</th>
<th>Advantages</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Interviews’</td>
<td>Data are collected orally. The interviewer asks clearly defined, systematic questions. Usually questions are predetermined and limited to a specific topic. Sometimes there are additional questions asked to elicit a more detailed response.</td>
<td>The data demonstrate specific examples of the observed outcome of the programmes. Interviews allow for flexibility.</td>
<td>The interviewer must be skilled in the process of interviewing and conduct the interviews in a systematic manner to ensure unbiased results.</td>
</tr>
<tr>
<td>‘Surveys’</td>
<td>The data are collected in a written format. Each respondent provides data on a set of clearly defined questions.</td>
<td>The data can be used to demonstrate specific examples of the observed outcome of programmes.</td>
<td>It is difficult to balance specific and general questions and ensure that larger or unexpected issues are not missed. Survey instruments must be completed consistently to avoid biased results.</td>
</tr>
<tr>
<td>‘Focus Groups’</td>
<td>A moderator guides a group discussion involving six to ten individuals representing specific stakeholders.</td>
<td>Focus groups provide specific, pertinent data. Group interaction can produce more information than individual interviews.</td>
<td>A specific set of skills is required of the focus group moderator. Data are difficult to summarise succinctly.</td>
</tr>
</tbody>
</table>
6.21 What Methods are Useful in Real-Time and Interim Evaluations?

Essentially, real time evaluation extends the level and detail of evaluation beyond the more routinised accumulation of data and information inherent in monitoring (see above). A detailed treatment of real-time evaluation is presented in chapter 4.

Real-time evaluation can utilise many of the methods used in ex post evaluations (see below), though it is of course unable to use data that can only be generated after the termination of the programme. Real-time evaluations have the big advantage, however, of being able to use observational techniques. The evaluators can participate in a wide range of programme activities, recording and documenting activities, getting a grip on the processes that are underway. The process of participation is liable to make the evaluation a piece of action research, in that the evaluators will themselves be intervening in the processes. Usually the terms of reference they are given will require that they do so, though it is possible to specify that a more detached position is adopted (at least for a period of time). Typically, the evaluators will report back on their conclusions as they are reached, and this will be used to improve programme performance in a continuous fashion. The main danger with such a stance is that the evaluators will become “captured” by the programme, feeling such loyalty to it that they are unable to make objective judgements. Exposure of their views and evidence base to outsiders can help avoid this. The danger with having “detached” and uncommunicative evaluators looking at an innovation programme is that participants may become increasingly suspicious of the ideas and intentions of silent observers.

Interim evaluations will generally be commissioned where there are some impacts that can be assessed, and thus will have more of the qualities of ex post evaluations.
6.22 What Methods are Useful in Ex Post Evaluation?

The ePUB RTD Evaluation Toolbox offers a useful overview of the range of methods and approaches that may be employed for ex post and real-time evaluation of programmes. (See illustration below, Box 6.7.) They span the entire range of methods discussed above, though the terminology employed is rather different from that used here.

The main feature of ex post evaluation is that we can expect to detect and assess impacts of the programme, even if we may sometimes have to examine intermediate rather than long-term impacts.
## Box 6.9 Methods employed in ex-post evaluation

### Quantitative Methods

- **Statistical data analysis**
  - *Innovation Surveys*: These provide basic data to describe the innovation process, which are summarised using descriptive statistics.
  - *Benchmarking studies*: These permit comparisons between sets of entities, based on a relevant set of indicators and accompanied by a reasoned explanation of their values.

- **Modelling approaches**
  - *Macroeconomic modelling and simulation approaches*: These are used to provide an assessment of the broader socio-economic impact of policy interventions.
  - *Microeconometric modelling*: These ranges of approaches permit the study of the effect of policy intervention at a variety of levels (individuals, institutes, firms, etc.). Mechanisms allow for the control of counterfactual events, through the specification of models that allow estimation of the effects on the outcome for participants had the programme not taken place.
  - *Productivity analysis approaches*: These allow an assessment of the impact of R&D on productivity growth at different levels of data aggregation. This set of approaches is particularly relevant for the analysis of the broader effects of R&D on the economy.
  - *Control group approaches*: These permit evaluators to capture the effect of programmes on different participants (and, in certain cases, non-participants) using statistically sophisticated techniques.

### Qualitative and semi-quantitative methods:

- **Interviews and case studies**: These approaches use direct observation of naturally occurring events, generally from the perspective of programme participants and stakeholders, to investigate behaviours, stimulated by the innovation programme, in their indigenous social setting.

- **Cost-benefit analysis**: Through the appraisal of all the economic and social effects of a programme, these approaches allow evaluators to ascertain whether a programme (or project) is economically efficient.

- **Expert Panels/Peer Review**: This type of approach is frequently used to measure scientific output, where they rely on the perception scientists have of the scientific contributions made by other peers. As such, peer review is the most widely used method for the evaluation of the output of scientific research. However, the approach has also been used for the evaluation of R&D collaboration programmes, such as Eureka.

- **Network Analysis**: Approaches of this type allow analysis of the structure of cooperation relationships and the consequences of participants’ decisions on actions. They provide explanations for the observed behaviours by analysing their social connections into networks.

- **Foresight/Technology Assessment**: Although more typically used for ex ante evaluation, these approaches may be used to identify potential mismatches in the strategic efficiency of projects and programmes.

Source: Polt and Rojo, 2002 – RTD Evaluation Toolbox
**Box 6.10 Methods in The Advanced Technology Program**

The ATP has, since 1990, provided R&D funding to US companies that have made proposals to pursue innovative technologies, responding to announcement of a competition for funding. Proposals are evaluated in technical and economic terms, with ATP selecting technologies that may be both high risk and high payoff, for many industries and applications. ATP has used a variety of *evaluation methods* to measure against mission, as depicted below.

<table>
<thead>
<tr>
<th>Method</th>
<th>Brief Description</th>
<th>Example of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytical/Conceptual Modelling</td>
<td>Investigating underlying concepts and developing models to better understand a program, project, or phenomenon</td>
<td>To describe conceptually the paths through which spillover effects may occur</td>
</tr>
<tr>
<td>Survey</td>
<td>Asking multiple parties a uniform set of questions for statistical analysis</td>
<td>To find out how many companies have licensed their newly developed technology to others</td>
</tr>
<tr>
<td>Case study - descriptive</td>
<td>Investigating in-depth a program, project, technology, or facility</td>
<td>To recount how a particular joint venture was formed, how the collaboration worked, and reasons for success – or lack thereof.</td>
</tr>
<tr>
<td>Case Study – economic estimation</td>
<td>Adding quantification of economic effects to a descriptive case study, using, for example, benefit-cost analysis</td>
<td>To estimate whether, and by how much, benefits of a project exceed its cost</td>
</tr>
<tr>
<td>Economic and statistical analysis</td>
<td>Using statistics and econometrics to analyse links between economic and social phenomenon, and to foresee economic effects</td>
<td>To determine how public funding affects private funding of research</td>
</tr>
<tr>
<td>Sociometric and social network analysis</td>
<td>Identifying and studying the structure of relationships to increase understanding of social/organisational behaviour and related economic outcomes</td>
<td>To learn how projects can be structured so that the diffusion of resulting knowledge can be increased</td>
</tr>
<tr>
<td>Bibliometrics</td>
<td>Tracking quantity of research outputs and assessing the frequency with which others cite publications or patents and noting who is doing the citing.</td>
<td>To find how many publications per research dollar a program generated and to learn extent and pattern of dissemination of a project’s publications and patents.</td>
</tr>
<tr>
<td>Historical tracing</td>
<td>Tracing forward from research to a future outcome, or backward from an outcome to contributing developments</td>
<td>To identify linkages between a public research project and significant later occurrences</td>
</tr>
<tr>
<td>Expert judgement</td>
<td>Using informed judgements to make assessments</td>
<td>To hypothesise the most likely first use of a new technology</td>
</tr>
</tbody>
</table>

6.23 What is the Case Study approach?

A Case Study is a more or less detailed examination of a particular instance of the phenomenon we are investigating. Often this will mean examining how a programme has affected one particular firm or organisation. (More often, there would be several case studies, exploring several different types of participant in, or target of, the programme.) If a programme has many specific projects or lines of activity, the case study may be one of these.

Case studies are typically undertaken to provide in-depth analysis of processes and outcomes. Surveys and other enquires across a large number of cases are good for establishing the broad distribution of impacts, and the statistical correlations that may exist between variables (e.g. firm size or project type, on the one hand, and degree of impact on the other). But we are not given direct information on the processes that give rise to these patterns, and one virtue of case study work is that it can help us examine processes. Often this involves reconstructing what happened over time in the organisation and its environment, and looking at what strategic decisions were taken. (Such decisions are usually critical to the extent of impact that involvement in a programme yields.)

This methodology is used in innovation programme evaluations to provide in-depth investigation of a particular aspect of the innovation programme and its reasons for success or failure. Here it must be noted that case study based evaluations involve far more of an activity than just the collection and analysis of information from the cases. Resources need to be allocated to analyse data and report it in ways that are useful to programme managers, staff, sponsors and any external stakeholders. The results need to be explored to understand what caused what and why and how much was attributable to the programme. Findings should be discussed and made available outside the programme, to policy bodies, industry councils and advisory groups. Peer review, including publication in journals could also be encouraged for wider dissemination purposes. Case study work is often a matter of participatory evaluation.

With the case study approach there is some evidence of a strong emphasis on the ‘learning’ aspect of evaluating innovation programmes (i.e. lessons drawn, training and knowledge exchange issues). See the illustration that accompanies this section.
Box 6.11 On Case Studies

A number of our project interviewees commented on this aspect with respect to using a case study approach to engage the user of the group as these are often the most likely people to make some limited or extended use of the evaluation results. This aspect is reflected in the following two comments from European based interviewees:

“The best way to see evaluation work is by interacting with those that would use the results of an innovation project to benefit their organisations. This can lead to some very clear observations. In a personal way, probably the skill (if that is the right word) most needed in evaluation is the ability to listen well and see points of view from different and divergent opinion”.

“Evaluations should be commissioned to the groups that undertake the job of running an Innovation programme. In my opinion bringing together a panel of experts to conduct the evaluation is not always the right tool for evaluation activities because you don’t have in-depth views about the programmes and the general situation the programmes have been operating under”. Therefore there should be more scope for in depth case study analysis type evaluations for innovation programmes”.
6.24 What sort of approaches are Participatory Monitoring and Evaluation?

“Participatory” here refers to the involvement of those who are participants in a programme, in the monitoring and evaluation of the programme itself. Practically all evaluations are liable to involve some participation from participants, of course, as informants, who provide information about the functioning of the programme. It is also good evaluation practice to seek comments from participants about the emerging conclusions of the evaluation. By exchanging information and stimulating dialogue, evaluation can contribute to continuous improvement – not only in top-down programme management, but more broadly in the stakeholder community. It is possible to involve participants much more fully than is common in evaluation and monitoring. Participatory approaches are tailored to the information needs of the participants in programmes, and uses them as major resources to determine the programme’s effectiveness. Participatory approaches have come to prominence especially in circumstances where community empowerment is of importance – mainly in development programmes and those aimed at deprived communities. Here the situation has typically been one of programmes being imposed by poorly-informed outsiders, and failing to achieve lasting benefits. Sometimes the programme goals have been inappropriate to local conditions; often the activities have petered away once the initial funding is removed. Participatory approaches should allow programmes to be more closely linked to the empirical circumstances, to engage the target communities more fully, and to provide that community with better information on which people can base their decisions.

The main problem with these approaches in the innovation environment is that the participants in innovation programmes are typically receiving substantial inputs of financial, consultancy, or other resources. They may be inclined to minimise problems and failings in the programmes if they feel that to do so is to threaten future largesse. They may see participatory approaches as offering a chance to promote their own specific agendas – to get their visions of future technologies and industrial activities prioritised. It is often seen that participatory approaches are liable to compromise the objectivity of the evaluation process. This has not stopped researchers and planners in development programmes from simultaneously developing ways to involve participants in the programme design, monitoring and evaluation, while seeking to make use of verifiable indicators and assessments. It is likely that some types of innovation programme could benefit from more use of participatory approaches.
Box 6.12 On Participatory Evaluation

Though fully-blown participatory evaluation approaches are rarely used in innovation programme evaluation, there is some awareness of the benefits of participation in the field. Thus one of the interviewees in this study remarked that:

“The most persuasive argument for using innovation programme evaluation is that if you make it a participatory process, then you should get more learning across the system, rather than it being limited to the immediate sponsors. All stakeholders need to have some involvement if at all possible” (From interview with UK Expert).

The principle of ‘participatory evaluation’ can be distinguished from more conventional, ‘traditional’ approaches to evaluation in that it has more emphasis on stakeholder participation. There may be multiple stakeholders who are or should be participants in evaluation, as they are directly or indirectly involved in deciding what a project or programme should achieve and how it should be achieved. They may include beneficiaries, project or programme staff and management, researchers, government agencies and any other financial donors. According to Estrella and Gaventa (1998) there are two main ways to characterise participation in monitoring and evaluation: by whom it is initiated and conducted, and whose perspectives are particularly emphasised. The first distinguishes between monitoring and evaluation that is externally led, internally led, or jointly-led. The second distinguishes between which stakeholders are emphasised – all major stakeholders, beneficiaries, or marginalised groups. The concept of learning is a key underlying principle of participatory evaluation. The emphasis is on ‘practical’ or ‘action-oriented’ learning. For instance, participatory evaluation can also be characterised as a process of individual and collective learning, like an educational experience for those various parties involved in the development of the programme. People become more aware and conscious of their strengths and weaknesses, their wider social realities, and their visions and perspectives of programme outcomes. It is this learning process that creates conditions conducive to change and action (Estrella and Gaventa, 1998).

Some of the useful online resources on participatory methods are:


6.25 Can we evaluate multiple programmes? The role of Cross cutting evaluations

Evaluation practice in the more advanced evaluation cultures seems to be moving less towards a focus on individual evaluations to examine individual programmes, and more concern with the picture that can be derived from a series of evaluations, or even from evaluations of a series of programmes. This is beginning to be documented in the study of “evaluation systems”.

We have been unable to find much documentary evidence, however, concerning cross-cutting evaluations, comparing the relative effectiveness of policies. One exception to this is the example of the National Audit Office (UK). Across innovation policy more generally, the UK National Audit Office and Department of Trade and Industry developed a methodology for comparing the comparative cost-effectiveness of the Department’s innovation schemes in delivering their objective of stimulating industrial innovation and increasing the competitiveness of UK industry (NAO, 1995).

“Grounded in decision theory, the basis of this approach was to identify 14 indicators representing effects on innovation and competitiveness and five representing costs. These were a mixture of process indicators (for example, average application processing time) and output indicators (for example, number of licences acquired). Values on each indicator were transformed onto ‘preference scales’ which scored each scheme on a scale of 100–0 on a relative basis, that is the best performing scheme on that indicator was assigned a score of 100 and the least performing zero, with the remainder interpolated. The process was repeated for the indicators themselves, to yield finally weighted preference scores based on equal weights for costs and objectives. By this methodology, the Teaching Company Scheme (a technology transfer scheme based upon joint supervision of young graduates by academia and industry) emerged as the most preferred scheme, and the LINK collaborative R&D programme the least preferred. Clearly, the outcomes are sensitive to the indicators selected and the weightings assigned to them. The low score for LINK was heavily influenced by poor performance on administrative indicators” (Georghiou, 1998).
Chapter 7 EVALUATING IMPACTS

7.1 What is Impact Assessment?

The demand for impact assessment can be seen as one element of the move to a knowledge-based society. Policymakers need to know what the results of their past policies have been, and to have a better idea of what the probable results of future policies will be. The idea of impact assessment is an obvious response to this demand.

Figure 7.1 Improvement of methods of impact assessment is vital to the success of innovation programme evaluation

There are many types of impact assessment, with several forms of *ex ante* assessment being very widely used – indeed having become mandatory in many institutions. Most prominent globally, it seems, is *environmental impact assessment*, whereby many planning decisions are subject to review in the light of their possible effects on, for instance, biodiversity, threatened species, carbon emissions, and so on.15 *Regulatory impact assessment* is common in the European Union institutions and member states, where the effort is being made to review policies before and after implementation – not least to limit the burden of regulations (especially on businesses).16

---

15 The International Association for Impact Assessment – [http://www.iaia.org](http://www.iaia.org) – is currently a particularly good starting point for work on environmental impact assessment, and also covers some work in social impact assessment.

16 The Communication from the Commission on Impact Assessment Brussels, 5.6.2002 COM (2002) outlines the process of impact assessment to be implemented in the Commission for all major initiatives. This notes that impact assessment is policy driven, focusing on examining “whether the impact of major policy proposals is sustainable and conforms to the principles of Better Regulation”.

Source: Supporting the Monitoring and Evaluation of Innovation Programmes survey data
There are many different traditions of impact assessment, with innovation programme evaluation being just one of these. While some topics can be addressed in a fairly experimental way – for example health and educational interventions can be piloted in a sample of hospitals or schools, respectively - many cannot. A number of challenges have to be confronted, and below we discuss the following major challenges to impact assessment in innovation programmes:

♦ **Complexity and Causality:** the difficulty to approach standard scientific and engineering practice in relating causes and effects and applying experimental methods.

♦ **Time Lags:** the long time that may be required before the desired impacts may be apparent.

♦ **Indirect Effects:** the likelihood that the impacts will be multiplicative (but harder to identify) as we move on from those directly involved in a programme to those who are in one way or another affected by the behaviour of the participants.

♦ **Multiple and Complex Objectives:** the range of impacts being considered may be a wide one, with many of the objectives being diffuse and hard to specify.
Box 7.1 Preliminary Impact Assessment Statement for European Initiatives

“The preliminary assessment gives a first overview of the problem identified, possible options and sectors affected. It will serve as a filter to help … identify the proposals that will be subject to an extended impact assessment”

1. PROBLEM IDENTIFICATION

Describe the problem that the policy/proposal is expected to tackle:
Indicate potentially unsustainable trends associated with the problem,
- Economically:
- Socially:
- Environmentally:
Indicate the potential inconsistencies between the three dimensions or with other policies

2. OBJECTIVE OF THE PROPOSAL

What is the overall policy objective in terms of expected impacts?

3. POLICY OPTIONS

What is the basic approach suggested to reach the objective?
What policy instruments have been considered?
In what way do the options identified respect the subsidiary and proportionality principles?
Which options can be excluded at this early stage?

4. IMPACTS – POSITIVE AND NEGATIVE

On a preliminary basis please indicate the expected positive and negative impacts of the selected options, particularly in terms of economic, social and environmental consequences
Please indicate who is affected and possible severe impacts on a particular social group, economic sector or region (inside or outside the EU), in the short term; in the medium and long term

5. FOLLOW-UP

What preparatory steps have already been taken (consultations, studies)?
Is an extended assessment recommended? Yes/no
Is a consultation planned? Yes/no

7.2 What Challenges does Impact Assessment face in terms of Problems of Complexity and Causality?

Social affairs of any kind are liable to be complex, and the effects of interventions difficult to assess. We shall consider here a number of facets of complexity, including multiple causality, time lags, indirect impacts and multiple objectives.

Rarely is it possible to conduct true experiments, or to control all major variables – even if these are well-understood. In some fields where there are many individuals whose experiences can be compared, and where outcomes are relatively easily contrasted, the problems may be less. The field of health technology assessment is an example where it can be possible to draw robust conclusions about the relative efficacy of different treatments, and even to estimate relative value-for-money (if we can put money values on to treatment outcomes). But innovation programmes deal with firms and other organisations, not just with individuals. They vary in ways that individuals do not, and the numbers of cases involved may be much lower.

In any case, it is fairly rare for us to be confronted with innovation programmes that offer true control groups, which would allow us to match cases of organisations that were or were not involved in the programme in the knowledge that they were assigned in a random fashion. More usually, those that are or are not associated with a programme differ in some way – they may be self-selected (perhaps more interested in innovation issues) or better-networked and thus more aware of the programme (perhaps more aware of innovative prospects in general), and so on. Comparisons between programme participants and non-participants will need to be made with caution – for example, matching them on as many variables as possible.

So it is difficult to employ robust statistical techniques in assessing programme impacts. But there are other difficulties, too. One is the problem of “over-determination”. By this we mean that the changes we observe in innovative performance may not, or not only, be related to the programme itself. They may be part of a wider complex of changes – say a growing focus on particular aspects of innovation – that has helped engender the programme and which would have resulted in some of the results that we might interpret as programme impacts. Furthermore, a programme is rarely a completely isolated phenomenon: there are often complementary policies and interventions from the same part of government and from other institutions.

The most important conclusion is that this set of problems has to be recognised for what it is: there are no complete solutions, and evaluation results will have to be interpreted with this in mind. We can do things like asking informants to estimate what might have happened if a programme was not instituted, or what effect the programme might have had without complementary changes. We can ask how it came to have the effects that it did. We can develop hypotheses about causality from some case studies, and test them in terms of other cases. We can see how far the accounts reached in the evaluation of a single programme are consistent with those reached by
other types of social enquiry (e.g. independent analyses of the innovation system, evaluation of other programmes\textsuperscript{17}). These are all standard approaches in social research more generally, which often has to struggle with the challenges of over-determination.

Finally, as with the other problems discussed here, the challenge is also an opportunity. For what the problems do is force us to examine more closely what is meant by innovation and what are the intended objectives of Innovation Programmes; and to understand more fully the innovation system within which interventions are being made. This understanding is vital for constructing a knowledge-based rationale for interventions.

\textsuperscript{17} It can be particularly useful to see if there is double-counting of impacts, in other words if the same changes are being attributed to more than one intervention.
7.3 What Challenges does Impact Assessment face in terms of Time Lags?

Another problematic issue concerns the long **time lags** that may be required for the impacts to be manifested. The goals of innovation policy are long-term ones, the economic pay-offs of individual innovations often require the development of markets and only slowly accrue sizeable revenues (perhaps over as much as 10-15 years). And of course many innovations do not succeed, not always because they are misjudged – sometimes unexpected products, standards, practices from other quarters have a negative impact. Often we cannot assess the full range of impacts in the short and medium term – indeed, this is rarely going to be the case, because of indirect impacts (see below).

Therefore, the best that can be hoped for in an evaluation will be to examine some leading indicators, or some early and possibly intermediate impacts that suggest that longer-term impacts are more rather than less likely. For instance, the ultimate intention may be to create a more effective innovation system, in which partners are sharing knowledge and collaborating widely and routinely to generate successful innovations. This long-term achievement is clearly more likely to emerge if the current programme has succeeded in getting the parties to work together and to establish ongoing linkages that are expected to persist beyond the programme’s lifetime. This sort of intermediate achievement is thus important to examine.

Intermediate impacts will often be those that are most clearly attributable to a particular programme, so in any case it is important to gather evidence on these if longer-term and more indirect “impacts” are to be plausibly related to programme activities.
7.4 What Challenges does Impact Assessment face in terms of Indirect Effects?

Indirect effects pose a related problem. Often an innovation’s real significance is not for the initial producers, but for its users: the economic gains made by the producer from selling a new product are a fraction of those made by users in employing this product to enhance process efficiency or the quality of their own products. There are upstream and downstream effects of innovation: a firm may demand changes in quantity and quality from its suppliers; it may open up new markets; or it may change practices in its customers. These changes may in turn prompt other parties with whom these suppliers and customers interact to modify their activities and behaviour. Tracing such impacts can be very difficult.

The most common answer is to use econometric or input-output models, which allow for some estimation of the knock-on effects of change at one point in a system, and where concepts of “embodied R&D”, for instance, can be employed. Such models often show, for example, that the overall effects of innovations on, say, employment, are quite different from those recorded at the level of the individual adopters of an innovation. Individual firms may shed jobs as they become more efficient. But overall market expansion may result in a net creation of jobs, in this case. These models are powerful tools, but are typically restricted only to fairly simple flows of products and changes in performance as recorded in financial terms. Improved collaboration and networking, changes in innovative practice, etc. are harder to assess in these ways, though modelling may be pushed forward to take greater account of less tangible factors.

It is also possible to trace impacts through a system by empirical means – for instance, looking at the diffusion of a specific innovation or practice through survey techniques, asking informants in specific locations about their knowledge and practices. One issue that arises here is that people are often very poorly informed about (and often not particularly concerned about) the ultimate sources of an innovative idea. Thus while many innovations have their roots in basic research, few firms consider institutions performing basic research to be of much interest to them. While this specific case is more of a problem in the context of R&D programme evaluation than in wider Innovation Programme evaluation, the point is a similar one. Directly asking people about whether a programme has influenced their behaviour may be a poor guide to understanding the indirect effects of a programme, because people are most aware of the direct influences on their own practice. The programme may be mediated through several steps before it affects a particular organisation, and tracing these steps is a laborious process, one which is certainly worth examining, but which is out of the scope of many evaluation studies.
7.5 What Challenges does Impact Assessment face in terms of Multiple and Complex Objectives?

The multiple objectives that characterise most Innovation Programmes – especially as we look across the range of activities they comprise – also pose a challenge for impact assessment. In some ways this is “merely” a practical issue – the more impacts that have been sought, the greater the number of indicators that will be required. Thus there may be problems of evaluation fatigue (if the same informants are expected to provide feedback on numerous indicators), of information overload (if the data cannot be readily synthesised and aggregated to form an uncomplicated account), and of workload on the part of the evaluators.

The solutions to these problems are a tactical matter. Among the approaches that may be adopted are:

♦ determining the key impacts on which to focus;
♦ determining whether some impacts can be assessed in some instances and others in separate evaluations, either later in time or of similar programmes;
♦ determining whether it is possible to examine certain indicators in depth in a limited sample of cases, and more indirectly or partially in others;
♦ preparing different reports on sub-programmes or on particular classes of impact.

Objectives can also be complex, in the sense that while it may be easy to state what the goal is, when examined in detail, the specific achievement desired is harder to elucidate. There may also be disagreement about what key terms and concepts actually mean. This is to be expected when aims such as “improved quality of life” are concerned, but it also applies to aims like “competitiveness”. In the management and economics literature there are different ideas of what this term means in practice, and what factors underpin it. Such different views are liable to imply a divergence of perspectives with regard to impact indicators and performance measurement. The evaluator needs to be clear about how concepts are being used, what approaches are being drawn upon - and how these may differ from those employed by sponsors and users! Ultimately, this derives from the fact that we are inevitably using one or other set of theories when attempting to study – and to influence – something like another innovation system. Enhanced understanding of the system may well require confronting different concepts of just what constitutes an effective system.
7.6 What sorts of impacts can be assessed, to what extent?

As the range of innovation programmes has expanded – for example, in terms both of the types of intervention covered (fiscal, knowledge transfer, network building, framework development, IPR regimes) and of the targets of such programmes (SMEs, start-ups, higher education establishments public sector research institutes, etc.) - so the range of impacts that might be evaluated has grown. Practically anything that can be expressed clearly and measured or verified, can be turned into an operational indicator – the problem is that obtaining information on some indicators can be very labour-intensive, and that some indicators are necessarily highly partial (i.e. only providing evidence on one part of the phenomenon being investigated – just as intelligence tests only measure certain facets of what is understood as “intelligence” in everyday discourse).

An example of the ways in which a broad objective can be translated into impacts that can be more precisely examined is provided by a study that set out to explore impacts of the ESPRIT programme on the competitiveness of participating firms (IFO et al., 2000). The evaluation put intangible assets at the core of competitiveness. The study used surveys of participants to examine how participation in the programme affected the ways in which participants created such intangible assets as knowledge, patents, brand names, reputation, networks and innovation routines. The evaluators considered how participation in the programme affected the ways in which participants used critical intangible resources such as R&D investment, technology acquisition. They examined impacts on business processes: whether new organisational forms were emerging around the innovations in question. Finally, more traditional aspects of competitiveness were considered, in terms of the “revealed” dimensions of outputs in terms of developing competitive products and services.

There is a need to exercise careful judgement in terms of the impacts that are selected as part of the evaluation process. A number of criteria may be applied in order to help prioritise the selection of these impacts. These are:

♦ What are the impacts of greatest relevance to the aims and objectives of the programme?
♦ What programme impacts are of greatest importance to the overall policy portfolio?
♦ Are similar impacts desired in other programmes where they may be more easily measured or where partial indications of their achievement may be combined with the results of the current evaluation?
♦ What are the relative costs of assessing the various desired impacts – are some easier to assess than others in terms of resources or methodological complexity?
♦ Can some impacts be substituted by proxy indicators or partial impacts?
Box 7.2 The socio-economic evaluation of the impacts of EUREKA

This study used a wide range of indicators to assess the potential impacts of EUREKA upon the competitiveness of participants. We cite it here to suggest the sorts of impact that may be assessed. Though the programme was mainly a research programme, it was aimed at having industrial impacts, and thus developed measures relevant to innovation programmes more widely. Questions were asked of participants concerning direct and indirect impacts, as outlined in the table below, where the figures relate to the proportion of participants reporting a “moderate” or “large”/ “very significant” impact.

<table>
<thead>
<tr>
<th>Direct effects</th>
<th>% reporting Moderate or large effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Product quality</td>
<td>46</td>
</tr>
<tr>
<td>- Co-operation with other companies</td>
<td>45</td>
</tr>
<tr>
<td>- Sales of products</td>
<td>42</td>
</tr>
<tr>
<td>- Market share</td>
<td>38</td>
</tr>
<tr>
<td>- Internationalisation of markets</td>
<td>32</td>
</tr>
<tr>
<td>- Co-operation with public sector R&amp;D organisations</td>
<td>32</td>
</tr>
<tr>
<td>- Sales of Know how or licences</td>
<td>22</td>
</tr>
<tr>
<td>- Number of employees</td>
<td>12</td>
</tr>
<tr>
<td>- Production costs</td>
<td>26</td>
</tr>
<tr>
<td>- Start-up of new companies</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indirect effects</th>
<th>% reporting Moderate or very significant effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Enhancement of firm’s knowledge base</td>
<td>76</td>
</tr>
<tr>
<td>- Leaking of Knowledge to collaborators</td>
<td>24</td>
</tr>
<tr>
<td>- Upgrading of personnel</td>
<td>59</td>
</tr>
<tr>
<td>- Recruitment of Personnel</td>
<td>23</td>
</tr>
<tr>
<td>- Know-how in management of collaboration</td>
<td>53</td>
</tr>
<tr>
<td>- Better integration between R&amp;D and other functions</td>
<td>36</td>
</tr>
</tbody>
</table>

Box 7.3 Start-ups from Universities as an Impact to be Assessed

The creation of university start-up companies is one particular type of innovation programme objective. Here the evaluation methodology indicators of interest are likely to be such parameters as the numbers of new start-ups and of jobs created. Germany provides a helpful example of one such innovation programme, EXIST, funded by the Federal Ministry of Education and Research since 1997.

EXIST intends to improve the climate for start-ups at universities, increasing the number of start-ups from academic institutions and enhancing the transfer of knowledge and technology from the universities. Models to motivate, train and support entrepreneurial personalities have been created in regional networks where the universities work together with such external partners as (extra-university) research institutions, companies, capital donors, technology and start-up centres, management consultancies, chambers of commerce, associations and local authorities. Together they develop an agreed offer for students, employees and graduates.

**Key outcomes**

There have been a huge number of activities in the regional networks like business plan competitions, idea competitions, seminars, lectures, events, assessment-centres, incubators, summer schools, international workshops etc. With EXIST 1076 start-ups were supported and more than 2600 jobs created.

**Evaluation of the programme**

The EXIST programme is monitored by the Fraunhofer Institute for Systems and Innovation Research (Fraunhofer ISI). The aim is to support the sponsor and the individual initiatives in implementing EXIST. The project support comprises at present three closely inter-related (work) foci:

- monitoring and assistance of the 5 EXIST and 10 EXIST-Transfer initiatives
- supporting them in general questions, as well as in exchanging experiences
- preparatory work for the sponsor (BMBF) and the project agency Jülich as well as analysis of further aspects in the work of the initiatives.

The scientific assistance is based on a process of mutual exchange and learning and is carried out in close collaboration between Fraunhofer ISI, the Federal German Ministry for Education and Research (BMBF), the project agency Jülich and the individual initiatives. Fraunhofer ISI sees its role as observer, adviser and discussion partner.

7.7 What is Additionality?

As evaluation practice has developed, and become more concerned with outputs and impacts of programmes and not just with the inputs they involve, so the concept of additionality has risen to the fore. Additionality is a deceptively simple idea. It refers to the change that can be attributed to the existence of the programme – what the additional effect of the programme is, as compared to what would have happened in the absence of the programme.\textsuperscript{18} Would the technologies have diffused anyway, would the innovation management practices have been adopted, were Universities and firms already on the way to establishing closer linkages? There is an understandable tendency for proponents of a programme to claim credit for positive changes that are part of a more general shift, rather than directly flowing from the interventions.

A further problem arises from the fact that, owing to the systemic nature of innovation and the desire to have interventions that have complementary effects, the effects measured as part of one programme evaluation may in fact be attributable, in part or in whole, to parallel programmes. Even harder to control are the broad level effects of macro-economic changes and economic cycles which, for example, may, in the event of an economic recession, militate against companies’ investments in R&D, or reduce employment rates at the same time a programme is being run to induce greater industrial investment in R&D or create jobs through the stimulation of company start-ups. In the latter case, however, such macro-economic effects are often clearly manifested and alternative indicators and impacts may be sought to provide evidence of the additionality of the programme.

The general problem with showing additionality is that, unlike in social or scientific experimentation it is often difficult or even impossible to operate a separate set of control conditions from which the effects of the intervention are isolated and to subsequently compare the two outcomes.

As with other aspects of evaluation, there are approaches to additionality that are more or less narrowly conceived. We can try to quantify the value-added of the programme in terms of such indicators as levels of technology adoption and economic impacts such as increased sales and employment, a higher share of profits coming from new products, a reduction in process costs or times, and so on. Such market-related effects form a traditional justification for programmes, and thus feature heavily in evaluations. Programmes with higher additionality are more likely to be contributing to policy goals (though it is pointed out by Polt (1998?) that this is not enough – additionality in the sense of positive returns to firms as a result of programmes may be a precondition for success, but private returns are not enough: the net social benefits of a programme must be positive).

\textsuperscript{18} See Buisseret, Cameron and Georghiou (1995) for a relevant discussion. See Georghiou’s chapter in OECD (1997b) for an insightful account of issues of behavioural additionality, which we draw on here.
7.8 What is Behavioural Additionality?

One lesson from research evaluation studies is that long-term impacts are much more than an immediate increase in sales of a new product or resources devoted to R&D or IPR protection. The issue at stake is not just a matter of more or less activity at one point in time. Activities have been reshaped, there has been a learning process in the individuals and organisations concerned. The real issue is whether a programme has effected a long-lasting change in innovation practices?

In other words, it is not that innovation would not have taken place anyway, albeit to a lesser extent. It is that the ways in which the innovation process is taking place have been transformed that are most significant. The innovations are coming more rapidly. They are more based, for example, on intelligence about emerging technologies and/or market requirements, rather than just being an incremental elaboration of what the firms were already doing. They involve more collaborative work and relate together more members of an industrial cluster. New skills have been acquired, new procedures and standards are in place.

The term “behavioural additionality” has come into use to describe the – hopefully enduring – changes in practices that participation in Innovation Programmes has induced. Evaluation studies need to be able to address behavioural additionality if they are to examine how far programmes are contributing to a reshaping of the innovation system.

Although we noted above that it is often not possible to provide control examples against which the effects of innovation programmes may be assessed, some limited opportunities exist. These include:

♦ The use of non-participants in a programme as a control group. Care must be exercised here to ensure that identical conditions prevail for the control and participant group – i.e. that the control group is not composed of failed applicants, or that they elected not to participate for reasons which would effect their behaviour vis a vis the desired objectives of the programme.

♦ In the case of innovation programmes that are restricted geographically or sectorally, comparison groups may be drawn from similar regions or sectors. Again care must be exercised to ensure that the two groups share similar characteristics and that spill-over effects from the programme do not occur.
Box 7.4 Some Examples of Behavioural Additionality

Georghiou and Claryss (2005) have recently prepared an OECD Working Paper. This presents a preliminary introduction and synthesis of work conducted to-date under the auspices of the OECD Working Party on Innovation and Technology Policy (TIP) to develop effective approaches to measuring behavioural additionality. The authors review the evidence on this topic, which has emerged from a series of national studies carried out in the context of the TIP working group during 2004-05.

“Behavioural additionality has generally been ignored by econometric studies of the effects of R&D support which focus on input additionality, where estimates are made of additional R&D expenditure or output additionality, whereby firm performance is compared between recipients and non-recipients of public support. These are both interesting questions, but in neither case is causality examined, nor is there an explicit or implicit model of how the firm uses public support. Such a model is integral to the concept of behavioural additionality. Behavioral additionality is directly related to an in depth insight into the dynamic capabilities which make a company different from its competitors.”

(Georghiou and Claryss, 2005, p.9)

Austrian and Irish studies provide examples of behavioural additionality methods:

**Austria:** Austria has combined a survey-based approach with an econometric analysis to quantify behavioural changes induced by funding from the FFF (Austria’s Federal R&D support scheme, its largest programme for financing business R&D). The survey included both recipients and non-recipients of FFF awards and compared responses regarding hypothetical scenarios (e.g. would you have cancelled the project if you had not received FFF support?) with those of actual scenarios (e.g. did you cancel the project after not receiving an award?). The econometric analysis, including around 1,100 R&D performers in time period 1995-2002, estimated the extent to which the stock of R&D personnel was affected by public money for R&D (Falk, 2002). The survey data revealed that FFF-funding was generating various dimensions of behavioural additionality:

- Around 80-85% of the sample firms experienced some degree of project additionality, i.e., an additional project was undertaken because of the grant.
- Acceleration additionalities arose for two in three firms, i.e., the project was brought to conclusion faster because of the additional resources made available to it.
- The share of companies appreciating scale additionalities ranged between 60-74%.

**Ireland:** In Ireland evaluations of two programmes, the ‘R&D Capability Grants’ and the ‘Competitive RTDI Schemes’ incorporated questions on behavioural additionality. The approach was to create a database of company information, files, and forms covering 700 companies. Interviews to survey company experience and views were carried out with 70 firms, including 10 non-participants in the schemes. The interviews were at the level of CEO and R&D Manager. In terms of additionality, 82% reported that the project would have happened without support (though only 22% would have done it in full, while 40% would have progressed at a slower pace and 14% on a smaller scale. A new issue in terms of additionality that was explored in this study concerned locational decisions, specifically whether the firm has been enabled to locate or keep R&D facilities in Ireland. Two thirds (65%) agreed with this proposition. Companies agreed that they were able to recruit higher skilled staff and to develop skills of existing staff. They also agreed that they had taken on higher risk R&D and around half agreed that the projects had made them change their business strategy or their long term manufacturing and business processes.
7.9 How does evaluation provide an insight to a systemic process?

The ‘systems of innovation’ concept has rapidly gained currency in both the academic field and in policy making circles. Historically, however, evaluation has tended to operate on a programme by programme basis, in isolation to the broader innovation policy portfolio. It is important to recognise that, by attempting to modify one part of the innovation system, policy interventions will initiate changes in other parts of the system, which may be beneficial or deleterious but rarely neutral in the longer term impacts. Thus, innovation programmes may serve to move the identified innovation “bottleneck” further downstream in the innovation process, or disturb the behaviour of actors in the innovation system, or even create new bottlenecks.

There are no hard and fast answers to this question. The VINNOVA Handbook identifies a number of challenges faced when attempting to evaluate the systemic effects of policy implementation: (IKED/VINNOVA, p. 42-42)

- **Policy mismatch and policy synergies**: here the question focuses on the externalities of the part of the system being evaluated. What other stakeholder, behavioural and structural factors come into play, either negatively or positively? Equally, how can the benefits accruing from the programme be enhanced by broadening its scope or by strengthening certain elements? Methodologies are needed to estimate the impact from other actors and institutions to increase the overall effectiveness of the innovation system.

- **Level of aggregation**: It is important to realise that the impact of an innovation programme may occur at a number of levels through the innovation system. Thus, if the evaluation is to capture these impacts, the choice of the level of aggregation for the evaluation (or indeed, evaluations) becomes critical. Indeed, it may be the case that the greatest impact is not felt by the participants in the programme. Similarly, programme impacts may be felt at various parts of a value chain outside of the industry actually targeted, because of competitive or other forms of interaction. Such additionalities must be captured and this may be done through aggregating the analysis and by investigating a larger number of actors or additional parts of the value chain, or by combining a number of complementary evaluations.

- **Risk**: The inherent risk associated with innovative activities means that some of the projects supported by a programme are highly likely to fail. Certainly, the level of success will vary among participants as will their expectations from the programme. For this reason, a large number of projects must be reviewed as part of a portfolio and the overall rates of success and impact examined. Here it is essential that the composition of the projects examined is appropriate to the specified objectives.

- **Dynamic effects**: As already stated the systemic nature of innovation implies that an intervention in one part will cause reactions in other parts of the system. Not only will these reactions be separated structurally, such as in other parts of the value chain, but they will also vary in the time they
take to occur. Some will be short term whilst others will take longer to become evident (for example, changes to fiscal and regulatory regimes). Much innovation policy aims to effect longer term changes in innovation systems. However, policy makers face pressure to determine the impact of their policies in the short term, thus inducing a demand to develop short term indicators of longer term effects.
7.10 What have we learned about innovation programme evaluation and the culture of evaluation?

This Guide, and the research in the SMIEP project, have suggested that innovation programme evaluation is an important and dynamic field of activity, but one that is still poorly developed in much of the EU and other countries.

Figure 7.1 Few countries have a well-established culture of evaluation in the innovation policy field

![Bar chart showing response to statement on evaluation culture.](image)

They have demonstrated that innovation programme evaluation is no simple matter, but that it can learn from the experience of other types of evaluation. There is no "magic bullet", but rather it is necessary to draw from a range of methods to get a grasp on necessarily complicated phenomena.

A major issue that has been repeatedly raised is the need to overcome the view of evaluation as purely a judgmental activity, rather than as an opportunity for learning. The view which has emerged through this Guide reflects the principles of advanced evaluation cultures:

- programmes are designed within sets of assumptions - evaluations can test these.
- planning for evaluation helps improve an overall understanding of the problems, the desired outcomes and the potential solutions, and makes evaluation itself more efficient and effective.
♠ programmes can always be improved (and evaluation is one source of insight for programme managers about how to achieve this).
♠ ‘failure’ contains important lessons - evaluation should provide these lessons rather than simply highlighting shortcomings.
♠ equally, ‘success’ cannot be reliably demonstrated without an evaluation.
♠ alternative modes of intervention may be available and possibly desirable - evaluations may identify these, and also pinpoint where there are contradictions and misalignments between different policy instruments.
♠ top down support for evaluation and acceptance of the results is required – which means informed users at senior levels.
♠ evaluation provides transparency on the operation of programmes to a wider audience and can help to engage stakeholders.
♠ evaluation offers an opportunity to benchmark programmes against each other and their forerunners/successors, and can allow for policy learning across different agencies and governments.
Chapter 8 REPORTING ON EVALUATION RESULTS

8.1 Why is it important to think about Reporting?

Reporting may be the last step in an evaluation – before its results are acted upon - but it is certainly not the least step. Indeed, the quality of reporting may have a considerable influence on just how far the results are heeded. And reports may be necessary during the course of an evaluation, not just at the end of the process. Finally, it is important to stress that report preparation should not be left to the end of the evaluation (see illustration).

Reporting presents the results of the evaluation study, together with the information necessary for readers to assess the validity and significance of these results – which will typically involve descriptive material on the programme and its context, and explication of the methods employed. Clearly these discussions can range from the terse and minimal to the discursive and lengthy. The exact point to aim at between these extremes will depend on the audience and the purpose of the particular report. Some different types of reports we can consider are:

Reports to the sponsor of the evaluation

♦ Interim reports during the course of the evaluation – typical in the case of extensive evaluations.

♦ Written final reports at the end of the evaluation process – practically every evaluation study will result in such a report. This will often be made publicly available after a process of consultation (e.g. programme participants are given a chance to correct errors)

♦ Verbal reports and presentations to accompany the above written reports. It is common to provide a succinct presentation on the final report; often there will be briefings as to the progress of an evaluation to accompany interim reports

Reports to Other Stakeholders

♦ Written reports for wider audiences – sometimes a shorter account is required when the full final report is seen as too long or complex for wide distribution.

♦ Reports and presentations to programme participants – often mainly a matter of tailored presentations based on the final report.

♦ Reports and presentations to other parties – for example, to policymakers and politicians, to other evaluation researchers, to innovation researchers.

♦ Evaluators can prepare ready made press releases that appropriate stakeholders can give to the media.
The content of the report – what elements of the evaluation are focused on, what the style of presentation is – needs to be based on the interests and capabilities of the audience. Their interests may involve, for example, the concerns of a specific audience with specific types of project or objective, or with particular industrial sectors, Universities, etc. Their capabilities have to be taken into account – for instance, the audience’s familiarity with jargon and techniques, whether the programme has already received a lot of exposure or whether the audience needs to be informed about it from scratch.
Box 8.1 NSF Guidance on Good Reporting of Evaluations

“Start early. Although we usually think about report writing as the final step in the evaluation, a good deal of the work can (and often does) take place before the data are collected. For example, a background section can often be developed using material from the original proposal. While some aspects of the methodology may deviate from the original proposal as the study progresses, most of the background information (e.g., nature of the problem, project goals) will remain the same throughout the evaluation. In addition, the evaluation study questions section can often be written using material that was developed for the evaluation design. The evaluation findings, conclusions, and recommendations generally need to wait for the end of the evaluation.

Because of the volume of written data that are collected on site, it is generally a good idea to organize study notes as soon after a site visit or interview as possible. These notes will often serve as a starting point for any individual case studies that might be included in the report. In addition... preparing written text soon after the data collection activity will help to classify and display the data and reduce the overall volume of narrative data that will eventually need to be summarized and reported at the end of the study. Finally, preparing sections of the findings chapter during the data collection phase allows researchers to generate preliminary conclusions or identify potential trends that can be confirmed or refuted by additional data collection activities.”

Source: chapter 7 “Reporting the Results of Mixed Method Evaluations” by Gary Silverstein and Laure Sharp in Frechtling & Sharp (1997)
8.2 How should reporting requirements be decided?

The reporting necessary for evaluation should be determined early on in the process. Indeed, the principles for this may well be established in the ROAME statement or other Programme Outline. At this point it should be accepted that an evaluation is not just the checking of a box to say that the programme was conducted properly, but is undertaken to ensure that lessons from the programme can be identified and disseminated. Dissemination plans should be part of the tendering and contracting of the evaluation.

On this basis, we can expect that the sponsor of the evaluation will have a considerable role to play in determining what sort of reporting will be required – whether to mandate an interim report, or simply a report at the end of the evaluation; whether to call for verbal as well as written presentations; whether to ask for specialised or summary reports for specific audiences; and so on. The sponsor may also go some way to specifying the structure of report(s) – e.g. issues that have to be considered, data that should be provided. What the sponsor cannot do is to specify what the conclusions should be!

There may well be scope for negotiation between sponsor and evaluator concerning precisely what reporting is to be done, and/or what reports should look like. If the evaluation study has been put out to tender among competing evaluators, they may be asked to make proposals about the form of reporting, for example.

There should always be a minimum expectation of what reports will be provided, and how these should be structured. Two more difficult matters may arise. First, it may become apparent that additional reporting should be undertaken: for instance, political or media interest in a programme or the issues it addresses may have grown rapidly. In such a case, there will be the need to consider who pays for the expenses of producing and disseminating reports, arranging or attending meetings. It is unreasonable to expect the evaluators to undertake a great deal of activity solely out of goodwill (or in the hope of gaining good publicity for their services). Contingency funds may need to be mobilised, possibly from the budget allocated for promotion of the programme itself.

Second, the evaluator may wish to make results – or other aspects of the study, such as methodology – more widely available than the sponsor originally foresaw. Wider dissemination of information on methods and results is generally beneficial. The main problems that can arise are: maintaining the anonymity of informants (see below); dealing with selective and even malicious media attention; and ensuring that the various parties in the programme have had a chance to correct errors and respond to specific criticisms. Generally, evaluations should be commissioned and conducted with the expectation that results will be made public.
8.3 How can we best present the results of an Evaluation to a particular audience?

While an evaluation report may be mainly targeted at one audience, it is likely to be read by a wider group. The type of audience will influence how much is already known about the programme and its background, how the evaluation results are to be used, and the amount of effort the reader is likely to put into assimilating the report. Some general points to bear in mind are:

♦ The sponsor of the evaluation and/or funder of the programme will want information on issues such as value-for-money and programme efficiency and effectiveness. Similar issues may well be of primary concern to other policymakers and politicians.

♦ Funders and sponsors will be more familiar with the programme structure and objectives, and this may need less reiteration for them. However, they may be more interested in a comparison of the effectiveness of this programme with others, and with the drawing out of lessons about future programmes.

♦ Participants in the programme, and those who might take part in future programmes, may be most interested in the benefits reported by people and organisations similar to them.

♦ Even those involved in the programme may be unfamiliar with the specialist jargon used by programme managers or in some other parts of the programme.

♦ Examples, case studies, and anecdotes may be particularly effective in communicating with lay audiences – but they are generally very useful tools.

♦ Graphs and charts are also compelling ways of presenting information, as long as there are not so many of them, or they are individually so complex, as to overload the reader.

♦ Only other evaluators will typically be interested in the technical details of the evaluation report, other than to assess the general validity of the results. Full details of methods can be presented in an Annexe.

♦ It will often be useful to prepare a full evaluation report for the sponsors of the evaluation, and to provide much shorter reports for other users.

♦ Indeed, a dissemination strategy for evaluation results is part and parcel of ensuring that the results will be used. However, the evaluators will need to agree this with the sponsors of the evaluation, and agreement in principle (and in writing) on such a strategy at a very early stage is desirable.
8.4 What does a report contain?

A short report may focus almost exclusively on the results of the evaluation, providing indications as to where the background to the study can be examined. The typical elements of a more extensive report are:

♦ **Executive Summary** – which, like the short report mentioned above, will typically focus on core conclusions, relating them to relevant sections of the main text.

♦ **Introduction**: setting the scene for the evaluation (usually this involves explaining who commissioned and undertook the evaluation, when and why).

♦ **Background** to the programme being evaluated, description of the rationale, objectives, and activities being evaluated.

♦ The **evaluation questions** – i.e. the sorts of impacts to be investigated, and the issues that these throw up.

♦ Brief overview of **methodology** – what information was generated, by what techniques.

♦ **Main results** – usually illustrated with graphs and charts.

♦ **Conclusions** – overview of major results and the implications that follow from them; recommendations for future programmes and policies (and evaluations).

♦ **Annexes**: more detailed notes on methodology and results.

While these elements will be part of almost all evaluation reports, they may be arranged in different ways. Reports may be organised along different lines. For example, a report may be organised thematically, in terms of different programme elements – or even in terms of specific themes that are being addressed. For example, an evaluation of a programme supporting research collaboration (COST) carried out in the late 1990s was intended to inform decisions about the future of the programme, including how it would be monitored and assessed on a continuous basis. After introductory explanation, the evaluation report was structured around five topics specified by an evaluation steering group (quality; structures, mechanisms and procedures; issues connected with widening the framework to incorporate central and eastern European countries; the programme as a catalyst for long-term international collaboration; and assessment of the investment in the activities by participants), and then discussed a series of issues concerning monitoring and assessment. (Here, recommendations were made concerning better documentation for monitoring purposes; improved linkages between different programme levels; improvements in the proposal selection process; procedures for set-up periods and mid-term evaluations; for prolongation and termination of actions; for committees, and for regular independent audit of cost finances, etc.) Finally, a set of scenarios were constructed, spelling out likely effects of different degrees of change in programme structures. (PREST et al, 1997)
**Box 8.2 Presentation of Evaluation Results: Discussion of Best Practices from the US Advanced Technologies Programme**

“Strategically Presenting Results.

Results have more effect if they are presented so that a nontechnical person can understand the science and commercialization. Results are presented in multiple ways—a brief abstract enabling someone to quickly grasp the key findings, an executive summary for someone who wants an overview of key highlights, and the full report. Quantitative findings are presented in tabular form, with graphics, and with accompanying qualitative analyses. We release many of our findings in fact sheets and make them available on ATP’s website. We have also published three special topic brochures that highlight projects in the health care, energy, and manufacturing sectors...

Another way that results and data are summarized is a ‘statistical abstract,’ an idea that we borrowed from the annual U.S. Census Bureau’s Statistical Abstract. Plans are to publish the ATP statistical abstract every other year—the first was released in September 2004, in a report called Measuring ATP Impact, 2004 Report on Economic Progress. The report describes the ATP using findings and data from recent reports and statistics. It also provides summaries of recent studies and ten detailed statistical tables that provide data on number and types of awards, technology areas, geographic regions, university participation, number of patents, commercialization, and post-award attraction of external funding. The data are presented for all projects and by characteristic of the project, by:

- Size of project lead,
- Type of project (single company or joint venture),
- Type of project,
- Technology area,
- Technology area by type of project.

…”

Source: Shipp et al (2005)
8.5 How can results be presented?

A report needs to strike the appropriate balance between accuracy and readability. It needs to be precise and detailed, without being too long or technical. It is important to engage and maintain reader attention. Some of the key issues that arise are:

♦ **Conciseness.** Evaluations studies typically generate large volumes of interview notes, survey returns, tabulations, and the like. It is necessary to be highly selective in using this material (especially in the main report: more detail can be presented in Annexes.) The bottom line is that information should be included that directly addresses the evaluation questions, and that contributes to establishing the overall arguments and conclusions reached. (This includes evidence that does not support these conclusions: such evidence has to be dealt with and accounted for, not suppressed.)

♦ **Readability.** Reports are often dry and technical affairs. Use of quotations and case study narratives can make the account more readable and memorable. There is a danger of too much narrative and quotation: the key rule here is to use these to illustrate points (for example, putting a relevant quotation on a text box next to an argument that is being made. Bullet point lists are similarly good ways of crystallising major points and breaking up dense analyses: and again they can be overused, leaving readers with a long series of poorly-connected points to somehow put together.

♦ **Illustrations.** These will usually be tables, graphs and charts derived from the results of surveys or archival analyses. (Even qualitative information can often be presented in tabular form. Topics such as “cases where informants reported that a particular outcome was achieved”; “views of informants as to which outcomes were most readily achieved” and the like, lend themselves to enumeration or citation of examples.) Again, there is a danger of overloading the text with illustrations: there is always a limit to the tables and charts that can be absorbed by readers, so decisions need to be taken as to which are most relevant instances. Other illustrations may also be employed – diagrams to represent organisational structure or task flow, possibly photographs of people at work in programme settings or of the sorts of technology being addressed, etc. The use of illustrations needs to be appropriate to the audience: readers must not feel talked down to, but visual design can be very important in encouraging readership.
8.6 What should be done by way of consultation and feedback?

It is common practice for a first draft of an evaluation report to be circulated to informants and those involved in the programme – they may receive a draft of the whole report or (often more usefully) just those parts dealing with their own fields of activity. Feedback from such readers is useful for correcting errors of fact and misinterpretations. Those whose opinions have been quoted, or who have recounted narratives that are being drawn on, may be able to point to and help to overcome misunderstandings and omissions in the account. (They may also point out if they feel their anonymity is being compromised.) People who may feel their performance has not been adequately represented are given a chance to respond to implicit and explicit criticisms.

While it is always possible that further feedback will be forthcoming after the report is published, early consultation is likely to allow the evaluators to determine where there is unhappiness with the report, and take steps to deal with this. The main danger is that the results of the evaluation will “leak” at this early stage, and that conclusions from a draft will be represented as the final word. Thus, when a draft report is circulated, this should be done in confidence, with those receiving it being asked not to distribute it further.
8.7 What should be done about confidentiality?

Maintaining confidentiality is liable to be an important issue. It is common practice to obtain views and comments from people who have been interviewed and surveyed on an anonymous basis. This allows them to speak in a forthright way, without being concerned that their comments to evaluators will lead to their being at the receiving end of angry telephone calls (or worse). Of course, anonymity should not be a cloak for rumours and half-truths to enter into the evaluation. Factual claims need to be checked as far as possible, and presented as being presented by one (or more) informants if they cannot be ascertained. Opinions and beliefs can be reported as such. The issue is, how far do these need to be tied to specific individuals?

Sometimes people will be happy to have their inputs attributed to them. But sometimes they will want specific things to be said in anonymity, and sometimes this will be the condition of an interview. Since it is difficult to negotiate levels of confidentiality in the case of surveys, these are often conducted on an all-or-nothing basis, and most often the respondents are informed that responses will be anonymous.

Usually the task then becomes one of (a) acknowledging the help provided by informants; (b) accurately representing the information that they have provided, which often requires some indication of what sort of respondent is involved this was (because this may well affect the way a reader will interpret the input); and (c) making it hard to associate a particular quotation or judgement with a specific named informant. It may be that the informant has been prepared to waive anonymity (point (c)), but this will not always be the case, and researchers have to be careful to avoid breaking any earlier commitments. (This of course implies that the evaluators will need to be very careful about any commitments that have been made in the first place.) It is often possible to preserve a fair degree of anonymity by attributing inputs to a respondent from a particular occupational or organisational background, but where there are very few people of this type this can be very problematic too. If the people acknowledged as having assisted in the evaluation only contains one individual from such a background, anonymity is compromised.

A related issue arises about “off the record” or “in confidence” remarks. These are sometimes provided by informants on the understanding that they will not be directly reported, but in order to provide the evaluator with necessary insight into the background of something being discussed. (Often this will be a matter of criticism of a specific individual or institution.) Such information can be invaluable, and usually it does not need to be reported in any direct way in order to be used to inform the evaluation report. If more explicit use needs to be made, it may be possible to anonymise the target of criticism or specific instance of practice that is being discussed. If this is not the case, the evaluator may need to negotiate with the informant concerning how the information can be represented.
Chapter 9 MAKING THE CASE FOR EVALUATION

9.1 How can we deal with Arguments against Evaluation?

Chapter 3 examined a number of arguments commonly mounted against evaluation, and suggested how these might be countered intellectually, and through practical measures that address the problems that they raise. We also outlined the arguments for evaluation, and have proceeded to discuss how evaluations need to be managed and implemented effectively. Ultimately, nothing is as convincing as examples of good evaluation practice, and of the results of evaluation studies making a real difference in policy practice.

However, this is only part of the story. It is one thing to rehearse arguments and document experiences, it is another to get people to listen to them. There are few if any short cuts here. It will be very valuable to win the ear of a senior decision-maker, who can exert influence on colleagues and subordinates. (Even then, there may be a token acknowledgement of evaluation from these parties, without any serious engagement with the issues raised. The terminology may be adopted and misapplied, raising the danger that an impoverished notion of evaluation becomes current.)

Exposure of evaluation principles and demonstration of their utility is very important – especially in circumstances where influential policymakers are present. They can demonstrate their approval, and alert subordinates should be able to draw their own conclusions from this. Peer influence is also a significant trigger for the development of favourable attitudes – accounts of what is being done in other countries can thus have a considerable effect. However, again, there can be problems. One of the – usually unstated – sources of opposition to evaluation approaches – especially specific techniques – is the “not invented here” phenomenon. If a particular policy approach is viewed as being closely associated with a particular country, political philosophy, or even one side of a local political argument, it may be rejected solely on grounds of association, without serious consideration of its underpinnings and potential usefulness in other contexts. Equally, as discussed below, there are dangers in uncritically copying approaches developed in one context by decision-makers in a quite different context – when to remain true to the underlying principles the approach would actually need to be modified to fit the new context. This is discussed in more detail below.
9.2 Who needs to be convinced? (1) champions

Two of the most outstanding and longstanding conclusions of innovation research are:

1. Innovations need product champions – senior figures who will fight for resources to be devoted to the innovation, even if it is still at early stages and a way from showing any profitable returns, and especially when the organisation is facing problems that lead other senior figures to suggest that only immediate activities should be prioritised.

2. Successful innovations are intimately linked to user requirements: the innovator has some way of understanding these requirements (perhaps from being a prime user him or herself, perhaps by continuous involvement with user communities – even through market research on occasion).

Rather similar issues arise when what we have to deal with is a policy innovation, such as introducing evaluation into an institutional setting, or introducing a new form of or role for evaluation activities.

A champion is required who will present the case to senior managers and policymakers, and fight the corner of evaluation when it comes up against opposition. The champion, usually a senior manager or political figure, will need to be armed with the arguments in support of evaluation. It helps considerably if he or she has an appreciation of the difficulties that evaluations can encounter, of the time scale involved in establishing and using results, of the challenges that confront using evaluations effectively. A sensitive manager will “sell” the evaluation principles to his or her subordinates, motivating them to engage fully with these, rather than just issuing an instruction to implement the process. Satisfactory implementation is liable to demand understanding and commitment to the process down the line.
9.3 Who needs to be convinced? (2) stakeholders

The product champion will have become a stakeholder, in the sense that his or her reputation may be tied up with the evaluation process. Other stakeholders also have to be won over to the process. By stakeholders we refer to all of those who have an interest in the evaluation. This can include, for instance:

♦ Those being evaluated (programme participants and managers), and those who might expect to be evaluated in future exercises.
♦ Possible users of those results concerning efficiency and value-for money – e.g. auditors and financial controllers.
♦ Possible users of the results concerning particular directions for innovation and innovation management, for instance research and technology communities who may be influenced by, generating and/or using the innovations in question.
♦ Possible users of more general results about innovation – its conditions and its consequences - including managers in firms that are or intend to be active in innovation or in the financing of innovations.
♦ Political figures and policymakers seeking to inform themselves about future directions for innovation policy and programmes.
♦ The media can become a stakeholder (when evaluations are relevant to ongoing stories or campaigns), and it will be important to provide clear exposition of the nature of evaluations to limit the extent to which results may be misrepresented.

As in the case of innovations, good linkage to users is not just a matter of telling users about the benefits of the innovation (in this case the policy innovation concerning evaluation). It also involves listening to users and hearing what they want. This does not mean necessarily giving users exactly what they think they want. Users are not always very good at seeing how they could be employing an innovation which they have not yet encountered, and they may be reluctant to change some or many of their practices. For instance, auditors will still need to continue performing basic auditing functions, even if they can be persuaded of the value of taking a broader look at things. But it does require taking account of what user capabilities for understanding evaluation results are, and what their needs for such results might be.

For these reasons, evaluation practices should be at a minimum designed so that the results can be presented in terms that are clearly and concisely explicable to non-expert users. In this context, some avant-garde quantitative results may be rather ineffective and while it is important to continue to experiment with fresh evaluation methodologies, this should not get in the way of presenting basic results. They should try as far as possible to engage with the sorts of programme impacts with which key users are concerned – this may be a focus on the competitiveness of regional clusters, on employment generation, on environmental impacts, or a host of other topics. The programme objectives will already specify key factors that the programme is intended to influence, but these also have to be related to the way in which such concerns are understood and expressed by key users.
9.4 What are major challenges that have to be overcome to win Support for the Use of Evaluations?

Clearly, the extent to which the institution has a well-developed culture of evaluation will be an important determinant of the use, and probably the quality, of individual evaluations. But there are factors that may render an individual evaluation less effective, even when it is in a generally conducive environment.\(^\text{19}\)

Among the challenges that most commonly have to be overcome are:

♦ Issues of Timing
♦ Resources allocated to evaluation
♦ Inappropriate tools and instruments
♦ Quality Control issues
♦ Independence issues

The SMEIP project interviews indicated that the first two challenges were most commonly experienced. Budget constraints and time limitations were repeatedly cited (sometimes within the same comment, such as a Scottish interviewee saying “Ultimately, our budgets are limited and the timescale within which things have to be done drive our choices”).

Other topics mentioned included:

♦ Overcomplicated regulations for the commissioning of evaluation (this affects civil servants' decisions about what programmes would be evaluated).
♦ Shortage of competent evaluators who are able to conduct impartial evaluation, transcend business models and use novel approaches. This was mentioned even in countries with a stronger evaluation culture.

\(^{19}\) EPEC (2005) usefully discusses factors that foster or discourage evaluations and their use at the level of individual evaluations, at the level of the institution more generally, and in policy processes.
9.5 How can we address problems of Timing?

Perhaps the most common problem is that evaluations do not fit well into the policy cycle. It takes time to conduct an adequate evaluation, but decisions may need to be made rapidly. Indeed, decisions about successor programmes, or about restructuring the whole innovation policy system, may need to be taken before a programme is complete. Hopefully the results of earlier evaluations can inform these decisions, but there may be pressure to share some early and provisional conclusions of an evaluation study in order to provide the specific intelligence that it is designed to yield. It will then be the responsibility of the evaluators to determine which conclusions may be released early, and how the necessary qualifications can be adequately reflected.

It is often hard to predict what timing issues will arise, because changing political circumstances may lead to demands arising for evaluation results at unexpected times. Similarly, they may mean that the moment has passed at which evaluation results could be most effectively used. Problems may arise, for instance, where there is a shake-up of senior political leadership, when a funding crisis happens, when there is sudden awareness of the need to intensively support a particular technology or cluster, and so on.

However, broad contours of the policy cycle can usually be predicted with reasonable accuracy. Funding rounds begin and end with some regularity; major reviews of policy areas are usually announced well in advance (as are elections); it is even possible to determine times at which decision makers are bound to be preoccupied with other urgent matters (even if it is harder to forecast when they will have a few free moments). Such issues of timing can be taken into account in the design and commissioning of evaluations. They may mean, for instance, that the evaluation should be started earlier rather than later, that it should be designed so as to produce key intermediate results at an early stage, that evaluators should be informed of the likelihood that their advice will be sought during the course of the evaluation, that monitoring arrangements should be well-developed and targeted at desired target outcomes, or that real-time or mid-term evaluations are conducted.

Equally, those sponsoring evaluations will need to clearly inform their political masters as to just what it is realistic to expect, and when. It is important not to raise false hopes about what a particular evaluation can contribute in the course of a particular policy cycle. It may be necessary to augment the partial results of an ongoing evaluation with information, evidence and conclusions drawn from earlier evaluations, monitoring activities and other sources.

It should also be noted that, in a fully developed culture of evaluation, the contribution of each successive programme evaluation should add more evidence to the overall picture of the national innovation system in an incremental fashion. Thus, at the higher policy (and political) levels a comparatively up to date assessment of the innovation environment is always
available, provided programmes have been launched at successive stages throughout the overall policy cycle.
9.6 How can we address problems of Resources?

It is commonplace to find statements along the lines of “evaluations typically cost 1 to 5% of the amount consumed by the programme being evaluated”, but this seems to be impressionistic rather than based on systematic survey. (It does correspond with the authors’ experience, however.)

Evaluations need to be sufficiently well-funded to carry out their mission. While evaluators may be motivated by non-financial reasons – such as curiosity concerning the policy area, the technologies involved, the innovation networks in play – these will not always be major factors, and there are inevitable costs of data production, interviewing, administration, report preparation and so on. Guy (1998) suggested that resources are encountered as a problem in many evaluation studies because they have not kept pace with the widening scope of such studies. Although, overall, more funds are now allocated to evaluation, those for individual evaluation studies have not grown: nevertheless the studies are expected to grapple with a wider and more challenging range of issues, to provide strategic advice and not just accountability, etc. Given that evaluations can contribute to staff training and development, it may be possible to think about funding some elements of the work from such unorthodox sources.

Evaluators themselves may have little say about the appropriate level of funding for a study should be, but they have to be realistic in communicating to sponsors about what they can achieve within the available resources, and should not promise too much. Otherwise, quality may be endangered. Programme designers and managers should regularly take soundings as to what resources may be required to undertake the sorts of evaluation they want, and not simply assume that the costs will be in line with those of earlier evaluations.

As with many other sorts of policy study, evaluation studies often suffer from a sort of “mission creep”, evaluators being expected to do more than earlier envisaged, to provide inputs and advice on a wider range of topics, and so on. Honesty is required in examining the funding implications of accumulating demands, and in saying just what can and cannot be done adequately within specific resource limits. However, if a Programme Model has been constructed, this should define the types of outcomes which the evaluation should investigate and also provide an assessment of the appropriate level of resources to do this.

Clearly, there is a strong case to establish a good level of dialogue between the programme management and the evaluators, whether the latter are internal or external to the institution or agency responsible for the programme. A clear shared understanding of the principal focus of the evaluation must be achieved and of the relative expectations of all the stakeholders.
9.7 How can we address problems associated with Inappropriate Tools and Instruments?

Several issues arise here. There can be a misleading focus on particular classes of data – for instance on financial or other quantified sorts of information, to the exclusion of others. While this may increase the likelihood of certain decision-makers accepting the evaluation study and recounting its conclusions, it is likely that this will not engender an appreciation of the complete and often complex results that the study provides. Decisions based on such a partial appreciation are liable to emphasise certain criteria rather than others, typically translating a broad policy goal into narrow and partial targets.

Another problem concerns inappropriate copying of evaluation techniques from one country (or policy area) to another. While general evaluation principles may sensibly be transferred, the way in which these principles need to be operationalised in different contexts will be very dependent upon the specificities of these contexts. Just as an Innovation Programme originally developed to boost innovation among a network of firms and other organisations in one country or economic sector or cluster cannot be expected to operate in precisely the same way in the different sorts of network that apply in another country, sector or cluster, so evaluation tools and instruments can be expected to vary.

Continuing with the example, an Innovation Programme designed for an innovation system in which there are weak links between technology suppliers and firms of a different sort may have to contain activities explicitly designed to foster such links, whereas another programme could expect to take them for granted and expect them to function so as to diffuse the practices embodied in the programme. The evaluations in both cases may be concerned with the flows of information and changes in behaviour mediated by the networks, but in the former case the evaluation study may need to address the question of whether new network linkages have been forged at all. And in practical terms, in the second case the evaluators may find that programme participants are well-informed about the networks within which they work, and can easily identify information sources and collaborators, while establishing this baseline information could be much harder in the first case.

The lesson is that evaluators cannot be expected to simply copy earlier studies and take their tools and techniques off the shelf, as it were. This may sometimes be possible – but the presumption when we are trying to learn from evaluation experiences in other countries (especially, but also from other sectors and clusters) is that much redesign of evaluation methods will be required.
9.8 How can we address problems of Quality Control?

It should go without saying that good quality evaluations are more liable to be used effectively than poor quality ones. But actually, this can only be taken for granted in well-developed cultures of evaluation. In situations where evaluation is less well established, there may be a lower capacity to assess whether an evaluation has been adequately commissioned, designed and implemented, and whether its results are being effectively used.

Programme managers and others who commission evaluations need to be informed clients. This means that they should themselves have received more than minimal briefing as to the utility and demands of the evaluation process. Ideally they will have had some training in commissioning and even in undertaking evaluations. They should thus be equipped not only to specify features of the evaluation process, but also to act as informed clients who can assess the quality and significance of any reports they receive.

Having independent evaluators (see below), using competition between evaluators as a way of assessing the quality of evaluation proposals, and being able to compare the results of different evaluations is important. Stimulating dialogue and discussion about evaluation approaches (for example through seminars and conferences, through exchange of experience across government departments and with peers in other countries) can contribute a great deal. Publishing evaluations, so that other parties can comment on their approach and results is vital.

Recently there have been efforts to develop quality standards for evaluations, which would mean that evaluators who aim to follow these standards would be providing some measure of quality assurance for their users. Examples of such approaches are illustrated below. While it might not be essential that any evaluators are committed to one particular quality approach, awareness of quality issues, and commitment to professional standards on the part of evaluators, is clearly essential.
9.9 Are there Standards for Evaluators to Follow?

Evaluators have been well aware that their work raises many ethical questions. The results of evaluation studies may not always be to the liking of all parties concerned – policies, programmes, individual projects may be deemed to have failed to meet expectations, for example. The results of evaluation studies are expected to influence future decisions, and so they have a practical significance well beyond that of most social research. It can be tempting to claim too much of an evaluation study, or to qualify critical remarks so heavily that their importance is obscured. Evaluators may find it hard to remain disinterested when their own future funding is under question, when they are liable to be criticised by one or other party.

For this reason there has been much interest in developing ethical standards for evaluation work. The illustration depicts one very elaborate set of guidelines for evaluators, set out in the United States. Some European evaluation societies’ standards are available online, at the following locations:

**France**
- English version
- French version
- German version

**Germany**
- English version
- German version
  [http://www.degeval.de/standards/Standards.pdf](http://www.degeval.de/standards/Standards.pdf)
- Russian version

**Italy**
- Italian version
  [http://www.valutazioneitaliana.it/deontologico.php](http://www.valutazioneitaliana.it/deontologico.php)

**Switzerland**
- English version
- German version
- French version

**United Kingdom**
- English version

### Box 9.1 The American Evaluation Association's GUIDING PRINCIPLES FOR EVALUATORS, July 2004

**A. Systematic Inquiry: Evaluators conduct systematic, data-based inquiries.**

1. To ensure the accuracy and credibility of the evaluative information they produce, evaluators should adhere to the highest technical standards appropriate to the methods they use.
2. Evaluators should explore with the client the shortcomings and strengths both of the various evaluation questions and the various approaches that might be used for answering those questions.
3. Evaluators should communicate their methods and approaches accurately and in sufficient detail to allow others to understand, interpret and critique their work. They should make clear the limitations of an evaluation and its results. Evaluators should discuss in a contextually appropriate way those values, assumptions, theories, methods, results, and analyses significantly affecting the interpretation of the evaluative findings. These statements apply to all aspects of the evaluation, from its initial conceptualization to the eventual use of findings.

**B. Competence: Evaluators provide competent performance to stakeholders.**

1. Evaluators should possess (or ensure that the evaluation team possesses) the education, abilities, skills and experience appropriate to undertake the tasks proposed in the evaluation.
2. To ensure recognition, accurate interpretation and respect for diversity, evaluators should ensure that the members of the evaluation team collectively demonstrate cultural competence. Cultural competence would be reflected in evaluators seeking awareness of their own culturally-based assumptions, their understanding of the worldviews of culturally-different participants and stakeholders in the evaluation, and the use of appropriate evaluation strategies and skills in working with culturally different groups. Diversity may be in terms of race, ethnicity, gender, religion, socio-economics, or other factors pertinent to the evaluation context.
3. Evaluators should practice within the limits of their professional training and competence, and should decline to conduct evaluations that fall substantially outside those limits. When declining the commission or request is not feasible or appropriate, evaluators should make clear any significant limitations on the evaluation that might result. Evaluators should make every effort to gain the competence directly or through the assistance of others who possess the required expertise.
4. Evaluators should continually seek to maintain and improve their competencies, in order to provide the highest level of performance in their evaluations. This continuing professional development might include formal coursework and workshops, self-study, evaluations of one's own practice, and working with other evaluators to learn from their skills and expertise.

**C. Integrity/Honesty: Evaluators display honesty and integrity in their own behavior, and attempt to ensure the honesty and integrity of the entire evaluation process.**

1. Evaluators should negotiate honestly with clients and relevant stakeholders concerning the costs, tasks to be undertaken, limitations of methodology, scope of results likely to be obtained, and uses of data resulting from a specific evaluation. It is primarily the evaluator's responsibility to initiate discussion and clarification of these matters, not the client's.
2. Before accepting an evaluation assignment, evaluators should disclose any roles or relationships they have that might pose a conflict of interest (or appearance of a conflict) with their role as an evaluator. If they proceed with the evaluation, the conflict(s) should be clearly articulated in reports of the evaluation results.

/cont.
### Box 9.1 continued

3. Evaluators should record all changes made in the originally negotiated project plans, and the reasons why the changes were made. If those changes would significantly affect the scope and likely results of the evaluation, the evaluator should inform the client and other important stakeholders in a timely fashion (barring good reason to the contrary, before proceeding with further work) of the changes and their likely impact.

4. Evaluators should be explicit about their own, their clients’, and other stakeholders’ interests and values concerning the conduct and outcomes of an evaluation.

5. Evaluators should not misrepresent their procedures, data or findings. Within reasonable limits, they should attempt to prevent or correct misuse of their work by others.

6. If evaluators determine that certain procedures or activities are likely to produce misleading evaluative information or conclusions, they have the responsibility to communicate their concerns and the reasons for them. If discussions with the client do not resolve these concerns, the evaluator should decline to conduct the evaluation. If declining the assignment is unfeasible or inappropriate, the evaluator should consult colleagues or relevant stakeholders about other proper ways to proceed. (Options might include discussions at a higher level, a dissenting cover letter or appendix, or refusal to sign the final document.)

7. Evaluators should disclose all sources of financial support for an evaluation, and the source of the request for the evaluation.

### D. Respect for People: Evaluators respect the security, dignity and self-worth of respondents, program participants, clients, and other evaluation stakeholders.

1. Evaluators should seek a comprehensive understanding of the important contextual elements of the evaluation. Contextual factors that may influence the results of a study include geographic location, timing, political and social climate, economic conditions, and other relevant activities in progress at the same time.

2. Evaluators should abide by current professional ethics, standards, and regulations regarding risks, harms, and burdens that might befall those participating in the evaluation; regarding informed consent for participation in evaluation; and regarding informing participants and clients about the scope and limits of confidentiality.

3. Because justified negative or critical conclusions from an evaluation must be explicitly stated, evaluations sometimes produce results that harm client or stakeholder interests. Under this circumstance, evaluators should seek to maximize the benefits and reduce any unnecessary harms that might occur, provided this will not compromise the integrity of the evaluation findings. Evaluators should carefully judge when the benefits from doing the evaluation or in performing certain evaluation procedures should be foregone because of the risks or harms. To the extent possible, these issues should be anticipated during the negotiation of the evaluation.

4. Knowing that evaluations may negatively affect the interests of some stakeholders, evaluators should conduct the evaluation and communicate its results in a way that clearly respects the stakeholders’ dignity and self-worth.

5. Where feasible, evaluators should attempt to foster social equity in evaluation, so that those who give to the evaluation may benefit in return. For example, evaluators should seek to ensure that those who bear the burdens of contributing data and incurring any risks do so willingly, and that they have full knowledge of and opportunity to obtain any benefits of the evaluation. Program participants should be informed that their eligibility to receive services does not hinge on their participation in the evaluation.

6. Evaluators have the responsibility to understand and respect differences among participants, such as differences in their culture, religion, gender, disability, age, sexual orientation and ethnicity, and to account for potential implications of these differences when planning, conducting, analysing, and reporting evaluations.

/cont.
Box 9.1 continued

E. Responsibilities for General and Public Welfare: Evaluators articulate and take into account the diversity of general and public interests and values that may be related to the evaluation.

1. When planning and reporting evaluations, evaluators should include relevant perspectives and interests of the full range of stakeholders.
2. Evaluators should consider not only the immediate operations and outcomes of whatever is being evaluated, but also its broad assumptions, implications and potential side effects.
3. Freedom of information is essential in a democracy. Evaluators should allow all relevant stakeholders access to evaluative information in forms that respect people and honor promises of confidentiality. Evaluators should actively disseminate information to stakeholders as resources allow. Communications that are tailored to a given stakeholder should include all results that may bear on interests of that stakeholder and refer to any other tailored communications to other stakeholders. In all cases, evaluators should strive to present results clearly and simply so that clients and other stakeholders can easily understand the evaluation process and results.
4. Evaluators should maintain a balance between client needs and other needs. Evaluators necessarily have a special relationship with the client who funds or requests the evaluation. By virtue of that relationship, evaluators must strive to meet legitimate client needs whenever it is feasible and appropriate to do so. However, that relationship can also place evaluators in difficult dilemmas when client interests conflict with other interests, or when client interests conflict with the obligation of evaluators for systematic inquiry, competence, integrity, and respect for people. In these cases, evaluators should explicitly identify and discuss the conflicts with the client and relevant stakeholders, resolve them when possible, determine whether continued work on the evaluation is advisable if the conflicts cannot be resolved, and make clear any significant limitations on the evaluation that might result if the conflict is not resolved.
5. Evaluators have obligations that encompass the public interest and good. These obligations are especially important when evaluators are supported by publicly-generated funds; but clear threats to the public good should never be ignored in any evaluation. Because the public interest and good are rarely the same as the interests of any particular group (including those of the client or funder), evaluators will usually have to go beyond analysis of particular stakeholder interests and consider the welfare of society as a whole.

Source: http://www.eval.org/Guiding%20Principles.htm
9.10 How can we address problems of Independence?

There are two facets to this issue. First, it is important that there is a pool of independent evaluators who can be trusted to accomplish high quality evaluations in an impartial manner. Second, it is necessary to strike a balance between specifying what sort of evaluation the users believe they optimally require, and the sorts of methodological practice that evaluators may consider particularly worth pursuing.

Independent evaluators may come from another part of the sponsoring ministry department or the government system, from consultancies, from Universities, and from other countries altogether (providing language and cultural barriers are surmountable, and there is not too steep a learning curve required to appreciate the specificities of innovation systems). Evaluation skills are becoming widely available as governments shift to more evidence-based policymaking, though it is not necessarily the case that those equipped to deal with evaluation in one policy area will appreciate the specific challenges of evaluation in a different field. Governments can play a role in stimulating the development of appropriate human resources, and in ensuring that there is dialogue between practitioners of evaluation in various fields.

A further danger is that even institutionally independent evaluators may, particularly if drawn from a limited pool of (national) expertise, not be truly independent. There may be an inferred or implicit pressure to deliver “positive” results regarding the performance of a programme in order to improve or maintain the chances that the evaluators’ services will be retained in future evaluations. Here, there are two issues. The first concerns the “honesty” of the programme managers in regarding the evaluation as a true opportunity for learning – and this forms another element of a true culture of evaluation. The second is that the onus is on the programme managers to explicitly state that there is no such thing as a ‘negative’ or ‘positive’ evaluation – merely a good evaluation which presents a clear picture of the programme, its outcomes and effects and contains valuable lessons for future policy making.

The quality of evaluations depends in part about how equipped the sponsor is to specify what is wanted from an evaluation. But there can be a tendency for sponsors to overspecify the approach that should be taken in an evaluation. While there are bound to be particular objectives and performance indicators that are essential to take into account, having been built into the programme from the outset, it is also important not to tie the hands of evaluators. At the very least they should be able to suggest additional issues and indicators that may need to be taken up; often the choice between evaluators will be a matter of the extent to which their proposals, in suggesting novel or interesting approaches or issues for examination, appeal to sponsors. Evaluators may wish to use particular methodologies, sometimes for poor reasons (for instance, this just happens to be the approach in which they specialise and they feel no need to develop other capabilities), sometimes for better ones
(e.g., they wish to test out new methods that promise to give better insights into the problem at hand).

The conclusion of this discussion is that the evaluators need to be:

♦ institutionally independent from those evaluated (programme managers and participants),

♦ practically independent – so that the evaluation can be conducted unimpeded, once agreement has been reached as to what the appropriate approach should be.

---

20 In the event that one part of an organisation is evaluating activities undertaken by another part of the same organisation, it is necessary to ensure that “Chinese Walls” are erected between the parties involved, to prevent undue influence being exercised on the evaluation.
9.11 What problems may arise in terms of unrealistic expectations of evaluation?

A major issue confronting evaluation studies (and many other policy analyses) is the need for realistic expectations, for those sponsoring and undertaking evaluations to “manage expectations” of exactly what is going to be achieved. As this Guide has made clear, what Innovation Programme evaluation can contribute is both more and less than is often anticipated by end-users.

Less, in that evaluations will rarely be condensable to a single indicator or even set of indicators, which gives an answer to the question of the value of a programme in a single figure or single sentence. Even when there is a resounding sentence at the outset of the evaluation report about whether the programme was well-justified or not, this will often be elaborated and nuanced as we read on! Less, because the evaluation will require the application of knowledge and judgement for its effective use, even if it does satisfy an immediate objective of having a legitimation of programme expenditure with which to confront senior officials. Expectations of what will be achieved thus need to be carefully managed, if the evaluation is not to be discounted, written off as over-sophisticated, or simply not used as effectively as possible.

More, because a serious and systematic evaluation should be providing a refreshing and hopefully timely overview of the field in question, placing the programme into the context of the evolving system which it is trying to influence. It should inform the design of future programmes and policies more generally, by providing explanations as well as descriptions. These explanations, crucially, should be telling us not just about what has worked or not worked in the past, but why this is so. This should give insights into the generalisability of approaches, into the possible impacts of quite new approaches or combinations of approaches, and so on.
9.12 How can we contribute to building an evaluation culture?

In an advanced evaluation culture, the evaluation process is appreciated as one important source of evidence for informing policy. So, evaluations themselves need to be conducted and presented effectively. They may be framed in terms of the aim of supporting improvement of programmes, and achieving increased value for money. (In doing this, they require better understanding of the linkage between activities, outcomes and impacts.) By charting the successes and failures of programmes, innovation programme evaluation helps create a clearer understanding of the problems to be addressed and the means to address them. It throws light on the nature of the innovation system, the ways in which innovation programmes are intervening in it, and can help to build realistic appreciation of what such programmes may achieve. Such awareness can be helpful not only to policymakers, since it can also achieve greater transparency of innovation programmes to all stakeholders – and broad understanding of the programmes and the role of evaluation is also part of building an evaluation culture.

It is important to conduct and present evaluations effectively, and to explicate why they are being undertaken and how they can be used. The evaluation culture is one with informed users, and the evaluation community has a role to play in helping users to inform themselves. To be sure, evaluation is more of a craft than a science – but it is not an occult art which only wise sages can undertake. Hopefully this Guide can contribute to demystifying evaluation, and providing users with some of the tools with which they can inform themselves.

In an advanced evaluation culture, informed users are not just passive recipients of the evaluation report. Opportunities should be taken to involve users in all stages of the activity – from Programme design and specification of the evaluation task, to providing feedback on the evaluation product itself. As well as making the evaluation task more effective, this will allow for much greater learning and policy impact.
9.13 Does it make sense to evaluate Programmes in isolation, or should an Evaluation Function be institutionalised?

Much is gained from having a system to commission, use, and overview individual evaluations.

1. Even in countries where evaluations are carried out regularly, there is quite often an “implementation gap” in that results from evaluations are not taken up at all or only in a “localised form” (i.e. they are only implemented if the institution evaluated could implement the recommendations on its own). This lack of feedback mechanism is probably the most important single factor limiting the value of evaluations. In many countries, there is a pressing need to secure the follow up on the appropriate level of policy making. This can be achieved in various ways. Some countries have put a formal obligation on those responsible for policy making to react to the results of evaluations (see Box 6.6 for the example of the UK ROAME(F) procedure). Others have a policy to expose the results of evaluations to a public discussion. A presumption in favour of publication of evaluation reports - though too much publicity might have its own drawbacks - is generally favourable both for the development of an evaluation culture as well as for encouraging policy implementation. The question of perspective is also useful here – evaluations of routine or small-scale programmes probably do not warrant too much exposure (who would be interested?) whilst those of national flagship programmes would elicit greater interest (and be open to a higher degree of risk).

2. When it comes to questions of the practical conduct of evaluations, experiences from several countries show that evaluations are best designed together with the programme or policy to be evaluated. Only such an early preparation can secure the collection and provision of the data needed in the course of the evaluation - especially where the use of micro-level data is essential for the estimation of the impacts. Best practice examples include the build-up of databases that can be used for various evaluations of programmes and policies addressing the innovative behaviour of a large number of firms as in the United States. Such a ‘concurrent design’ of both programme and evaluation procedure would also raise the common acceptance of the procedures and criteria among the persons and institutions involved.

3. Rather than being conducted on an ad-hoc basis, evaluations should be carried out frequently. There are substantial learning effects of frequent evaluations, both on the demand side of policy makers, programme managers and providers as well as on the supply side of the professionals carrying out evaluations. Knowledge accumulated from evaluations could be a valuable source of policy making, as the practice in Nordic countries exemplifies. Frequent evaluations are also shown to have a lasting effect on the behaviour of the evaluated institutions and in many cases contributed per se to the improvement of routines and performance.
9.14 How can we affect how far the results of Innovation Programme Evaluations are applied?

Getting the results of evaluations applied is more than a matter of fine-tuning programmes. Evaluation can help to enhance learning within the policymaking environment. A number of implications flow from this. Though wide dissemination is important (as long as it is not overly resource intensive), it is important to get the (broad) results of evaluation to the people that matter.

Evaluation reports should make specific recommendations as to actions that should follow the report. Design of reports so that this is accomplished in a lucid, reader-friendly fashion, is vital. Establishing an understanding of the analysis that underpins the recommendations – i.e. precisely which results it is built on – is also vital. The recommendations have to be argued for and clearly based on the evidence obtained, so that users can in turn argue for them (with line-managers, policy makers, financial authorities, politicians, etc.), and not have to just take them on faith as stemming from wise experts.

But ensuring the use of evaluations involves more than just writing good reports, helping to disseminate them, and making persuasive presentations, important though all these are. It involves preparation before the reporting stage, and follow-up after it.

Those sponsoring the evaluation should have established a timetable and set of procedures to govern the reporting and dissemination of the evaluation results and a means of assessing the use that is being made of them. Ideally, this will be embedded in the Programme Outline: it is critical in the request for evaluation proposals.

During the evaluation process, every opportunity should be taken to inform those contacted about the scope and goals of the evaluation, the ways in which its eventual results might be used by the various parties concerned, and how they can get hold of reports. Interest in these matters should be recorded so that the most engaged parties can be invited to dissemination events, and perhaps to comment on drafts of reports. (These activities can be expensive, and costs will be a limiting factor.)

An implementation plan allows for the checking of progress against recommendations. Even if such a plan has not been formally prepared, it should be possible to determine which agencies should be responsible for following up which recommendations, and which parties (in addition to the programme manager and/or sponsor of the evaluation) should be responsible for checking on progress. While it may seem onerous to spell such issues out in detail, having a framework to examine implementation is much more of a basis for progress than is simply “muddling through”. For one thing, such a framework may identify actions that were not clearly thought of before, and indicate where steps might be required earlier rather than later.
SMART INNOVATION – Supporting the Monitoring and Evaluation of Innovation Programmes

BIBLIOGRAPHY

(Texts which are particularly useful, but are not discussed in detail in the main text, are briefly outlined in footnotes.)


Bradbury, M and Davies M The Evaluation of Support Programmes: The Example of the United Kingdom, OECD STI Review No. 21 1998


21 This describes how the Dutch government chose not to undertake its own evaluations of research institutes and groups; instead it aimed to stimulate these organisations to perform or commission their own’ evaluations and to use these to formulate strategies.
22 This review of the use of evaluation in EC policies and programmes covers a wide range of policy fields, and only incidentally touches on innovation policies. Nevertheless, it carries a great deal of evidence about the use of evaluations in the policy process of the EC, and draws valuable lessons – based on careful study of various users of evaluations – that are relevant to the innovation policy field.
SMART INNOVATION – Supporting the Monitoring and Evaluation of Innovation Programmes


Patton, M Q. *Qualitative Evaluation and Research Methods* (2nd ed) Beverly Hill, Sage 1990


23 Report overviewing country experiences in the evaluation of basic research, but containing much insight into evaluation cultures in different countries, problems of quantitative and qualitative approaches, and other relevant topics.


