Evaluating Foresight and Lessons for Its Future Impact

Luke Georghiou
PREST, University of Manchester[1], UK

Introduction

In the array of strategic policy approaches available to the policymaker, evaluation and foresight are often grouped together, with the simple distinction being made that evaluation is looking backwards at what has occurred and foresight looking forward at possible futures. The two approaches are not so easily separated as most evaluations include some form of formative perspective on the future and often have to consider the future implications of the measures they seek to assess. In turn, foresight activity generally needs to be informed by a thorough understanding of the past and present. However, the object of this paper is to turn one instrument upon the other – to examine how we might go about evaluating foresight and what conclusions that might lead us to in terms of the future of foresight.

In many ways the time is right for an evaluative approach. Technology foresight at a national level may now be seen as a policy instrument which is approaching maturity. Since the early 1990s the practice has diffused widely to the point that most industrialised countries and several advanced developing countries have experience of some form of foresight exercise. Many have been through more than one iteration, while others are about to do so. Despite this spread of experience there has not so far been a serious effort made to understand the effects of the wave of foresight activity. In particular, foresight has not been systematically evaluated as an instrument of science and innovation policy. In the following sections, after some consideration of what we mean by foresight, the problems and issues involved in its evaluation are discussed. Some experiences of efforts to evaluate individual programmes are considered and some emerging lessons for foresight as a policy tool are discussed.

As a starting point in assessing the international experience, it is important to stress not only what is common to foresight activities but also the ways in which they differ. In reality, what we term foresight covers multiple activities and purposes sharing a name. In terms of purpose, some common goals for foresight are:

- *Exploring future opportunities so as to set priorities for investment in science and innovation activities*. The degree to which priorities can emerge from foresight varies from "critical technologies" exercises where the whole discourse is focussed on a priority list, through more general programmes from which priorities are derived, to targeted foresight where the priorities are in effect set

[1] Thanks are due to Michael Keenan for some very helpful suggestions and for co-developing with me the perspective on evaluation of foresight presented here.

before foresight begins. As we shall discuss below, the real effect of foresight on priorities may be difficult to determine.

- *Reorienting the Science & Innovation System.* This goal is related to priority setting but goes further. In such cases, there may have been a preliminary diagnosis that the science and innovation system does not match the needs of the country. This was a common situation in Central and Eastern Europe in the immediate post-Communist period when, apart from severe resource difficulties, capabilities reflected an industrial system that no longer existed. Foresight has been used as a tool to re-orientate away from fields such as materials research and towards life sciences as well as to explore new institutional structures.
- *Demonstrating the vitality of the Science and Innovation System.* In this context foresight becomes a "shop window" to demonstrate the technological opportunities that are available and to assess the capability of science and industry to fulfil that promise.
- *Bringing new actors into the strategic debate.* A growing tendency is the use of foresight as an instrument to broaden the range of actors engaged in science and innovation policy. One example is the inclusion of social stakeholders or even sections of the general public such as youth.
- *Building new networks and linkages across fields, sectors & markets or around problems.* A different type of reorientation is sought when foresight is explicitly aimed at creating new networks and or clusters which break out of long-standing disciplinary or sectoral ties.

The modalities of foresight may also differ strongly. All of the above goals may be pursued at organisational, local, regional, national or supranational levels. The timescale of foresight ranges from the immediate future to the far horizon. The range of actors involved, the process and methods used, and even the status of the activity varies considerably. Foresight ranges from methodological experiment through to major politically driven initiatives.

Against this background of variation, one of the more flexible definitions available will be used to define what is discussed below. The origin of the definition is itself evidence of the cumulative learning effects involved in the international diffusion of foresight, having been formulated as part of the efforts of one of the most recent (and smallest) countries to engage in this activity. It captures key elements of the process that are usually neglected in some of the more commonly used formulations:

> "The foresight process involves intense iterative periods of open reflection, networking, consultation and discussion, leading to the **joint refining of future visions** and the **common ownership of strategies**, with the aim of exploiting long term opportunities opened up through the impact of science, technology and innovation on society….It is **the discovery of a common space for open thinking on the future** and the **incubation of strategic approaches**…"[2]

Of particular importance here is the stress placed upon the way in which joint foresight activities are linked to the joint formulation and ownership of strategies. This

[2] Cassingena Harper J (ed), Vision Document, eFORESEE Malta ICT and Knowledge Futures Pilot (January 2003)

perspective avoids what it will be argued below is a common trap, the treatment of foresight and its implementation as separate processes without serious attempts to build bridges between or to link the two.

Evaluation of Foresight

Since foresight is a policy instrument consuming time and resources, it is reasonable to expect that it should be subject to evaluation of a comparable rigour to other tools. In a generalised evaluation framework, three basic tests could be applied:

- Accountability – with questions such as whether the activity was efficiently conducted and proper use made of public funds;
- Justification – with questions such as whether the effects of foresight justify its continuation and extension; and
- Learning – asking how can foresight be done better in particular circumstances.

In a standard evaluation approach, it is important to define the scope and purpose of what is being evaluated at an early stage[3]. The variety of forms of foresight has been discussed in the previous section. Another dimension in which foresight has to be lineated is that of location in time. The key question is where does a foresight activity begin and where does it end. In a first national effort, the beginning is usually clear as the process is initiated with a decision to commit resources and often to establish some sort of secretariat. The end is frequently much less clearly delineated. Where the aim is a report or list of priorities, publication and launch marks some kind of termination though dissemination and other implementation activities may well follow. The launch of networking activities is far less likely to offer a clean break as these are likely to persist for some time after the foresight activity has ended. An arbitrary decision may need to be made on when to demarcate the cut-off point by when foresight outputs cease to provide a distinct or influential voice in policy discourse. Furthermore, to understand the context in which foresight is operating it is necessary to locate it a broader strategic and policy context. The evaluation will have to explore the period in which foresight emerged and its interaction with other elements of the system.

The timing issue is also linked to the type of question being asked. If a linear or sequential view of foresight is taken, process issues are best pursued while the activity is still under way. However, many outputs and outcomes will not be clearly visible at this time and will need to be investigated ex post. Here the problem becomes one of attributing effects. If accuracy of the future visions is an issue the ex post delay corresponds to the foresight period. This may not be a problem with short horizon five-year critical technology exercises, but it requires a remarkable stable system if the issue is to be usefully pursued for foresight on, say, a fifteen-year timescale. Only the STA/NISTEP forecasts have been properly assessed on this basis.

[3] Gibbons M and Georghiou L, Evaluation of Research – a Selection of Current Practices, OECD, 1986

In my paper for the previous NISTEP International Foresight Conference[4] I posited a generational model of foresight which, in brief, distinguished first generation activity concerned with technological forecasting by experts, second generation bringing in industry and the market and third generation foresight adding a social and user-oriented perspective. Implicit in these models were different approaches to evaluation. For first generation the key issues are accuracy of prediction and diffusion of results (to non-experts). In the second generation the take-up of priorities and establishment of networks become key evaluation issues, while the third generation implies the involvement of stakeholders in evaluation and looks for evidence of the mergence of a foresight culture. It should be stressed that these generations are ideal types and that an individual foresight activity may exhibit elements of two or even three generations. It does however, emphasise that the approach to evaluation is conditioned by the approach to foresight.

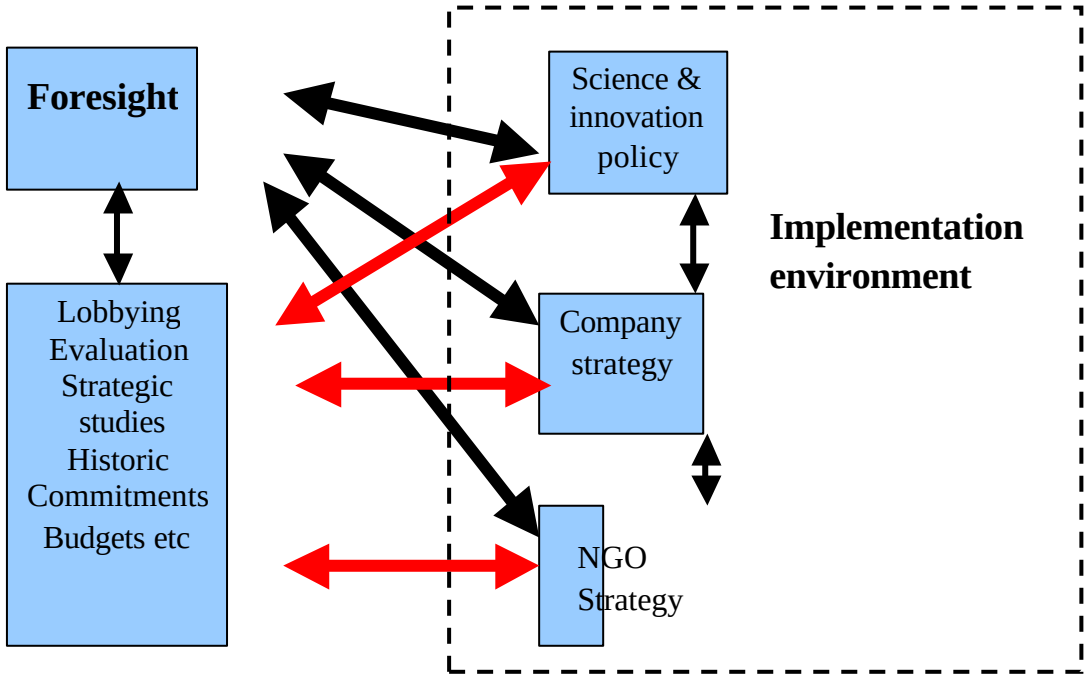
Process evaluation covers topics such as organisation and management, and would for example ask: Were the 'right' people involved? Did expert panels receive adequate support? Was the exercise adequately linked to decision-making centres? It may also address the question of the appropriateness and efficiency of methods used, for example: Should a Delphi have been used? Were scenario workshops properly facilitated? As noted above it should be conducted in real-time or immediately after an activity is complete to ensure that the findings are not distorted by hindsight or obscured by loss of data.

Moving to outputs and outcomes, probably the most important observation is that outputs measure only activity and not its significance. Hence it is useful to know numbers participating in meetings or surveys, reports disseminated, meetings held, website hits and so on but none of these measure the effects of these contacts or their contribution to outcomes. Numbers may even be misleading; the number of "new networks" formed disguises variation in their novelty, size, significance, durability and so on.

Figure 1 is intended to illustrate that assessing the effects of foresight requires an understanding that it is only one of several influences upon public policy or the strategy of firms. Furthermore, once a foresight output enters the implementation environment the question may be asked of how is it different from other policy information? Possibly the answer lies in a longer timescale, creativity or commitment but all of these elements can also come from other sources. The implication is that evaluation of foresight must include understanding of the interaction of foresight outputs with the strategic behaviour of policy and economic actors.

[4] Georghiou L Third Generation Foresight – Integrating the Socio-economic Dimension, in Proceedings of the International Conference on Technology Foresight – the approach to and potential for New technology Foresight, NISTEP Research Material 77, March 2001

Figure 1: Foresight in a Non-Linear Relationship with its Implementation Environment



There are also some normative issues involved. Foresight is not always tuned to the needs of recipients and hence, to extend the analogy, the signal may be obscured by noise and not picked up. Information needs to be presented in such a way that policy/strategy mechanisms can receive and absorb it. One moderating factor is that of timing. This needs to synchronise with policy and strategic cycles[5]. Furthermore, the level of recommendations needs to match available funding or capacity for reform. However, foresight cannot always work within the status quo and occasionally it is the policy/strategy structure that needs to change in the light of foresight information.

A key question in the evaluation of any public policy intervention is that of additionality – the extent to which the activity would have taken place without a public intervention. In this framework, the questions which should be asked about a foresight activity are:

- Would foresight have happened without the policy intervention?
- Is foresight done differently/better because of the policy intervention
- Are the resulting actions better because of foresight
- Have persistent changes been achieved (eg foresight culture)?

Within the field of evaluation, recent thinking has moved away from treating additionality as a binary stop-go item. Within a systems of innovation framework, temporary financial interventions are seen as less important than efforts to change the

[5] See the doctoral thesis of Michael Keenan An Evaluation of the Implementation of the UK technology Foresight Programme, PREST, University of Manchester 2000 for examples of this type of implementation barrier in the first UK Foresight Programme.

innovation system for the better in a lasting way. If it is accepted that foresight is correcting an inherent tendency to have excessively short-term horizons and a difficulty on forming new networks around technologically and socially innovative activities, then foresight may be best evaluated ultimately in terms of its ability to change values and behaviour in these directions[6].

The evaluation of foresight should also beware of potential traps. Traditionally foresight is seen as a process of building commitment among stakeholders –an important element for example in Martin's "5Cs". However, from an evaluation perspective this also creates risks when trying to assess the additionality of foresight. One risk is that of the self-fulfilling prophecy when the "owners" of a foresight activity (for example a sponsor ministry) also control the distribution of resources at the implementation phase. There may be a tendency in this situation to cause foresight priorities to have a stronger influence in the implementation environment than may be justified in terms of the rigour and merit of the exercise. At a more methodological level, stakeholding and consensus may be seen to some extent as a trade-off with creativity and insight. It may be somewhat easier to get "buy-in" to a set of views that are already commonly held than for a really novel or disruptive idea.

Some Evaluation Experiences

Table 1 shows some recent experiences of foresight evaluation, and serves mainly to emphasise that a consistent and comparable approach has not emerged. Two of the countries mentioned can be discussed further as cases studies.

Table 1: Evaluation of National Foresight Activities

Country	Evaluation Effort
Austria	Internal assessment of impacts by Science Ministry
Netherlands (OCV)	Self-evaluation, PhD study, Masters thesis, evaluation by Advisory Council for Science & Technology (AWT)
Sweden	Process (and not the impacts) evaluated continuously by an Evaluation Committee
Japan	Assessment of realisation of results some 15-20 years after identification in STA forecasts
Germany	Delphi 98 evaluation questionnaire; FUTUR evaluated during 2002
United Kingdom	Grand plans that degenerated into piecemeal efforts; some limited external (and independent) scrutiny, e.g. by Parliament, a PhD study, etc.

[6] Georghiou L, Impact and Additionality of Innovation Policy, in Boekholt P (ed) Innovation Policy and Sustainable Development: Can Innovation Incentives make a Difference, Brussels: IWT-Observatory 2002

Case 1: UK Foresight Evaluation Experiences

The first case, that of the UK, illustrates that without a consistent, credible central approach to evaluation, the likely result is a proliferation of activity. Much of the work listed in Table 2 was at a sub-critical level, or else relied very heavily on anecdotal and potentially prejudiced evidence. It may be seen that the operating Ministry for foresight, the Office of Science and Technology was the main driver of activity but, despite commissioning a number of methodological studies and some fieldwork, it never allowed these to be compiled as an authoritative evaluation. Other interested bodies such as the Parliamentary Office of Science and Technology were able to put forward more critical and insightful views but lacked the resources to follow up in terms of extensive collection of evidence. A parallel theme was growing government enthusiasm for performance indicators in all aspects of public sector activity. Notwithstanding the comments made above about the limitations of output indicators the constant pressure was to capture the effects of foresight in terms of key indicators. During the Second Cycle the author was one of those asked to develop an indicator driven evaluation framework. A particular difficulty with this accountability-style of approach is that foresight depends heavily upon the unpaid involvement of panellists and other contributors who do not take kindly to being monitored. A "softer" evaluation approach was adopted which relied upon participants to collect and analyse a significant part of the data, while other items would be compiled centrally. The organising principle was to separate process from impact and in the latter case to identify the five main stakeholder groups: the science base, industry and commerce, the voluntary sector, government, and education, training and public understanding of science. From the framework a set of key indicators was derived.

Table 2: UK Evaluations Relating to Foresight

- OST/PREST conducted survey of panellists (1995)
- OST drafted more comprehensive evaluation proposals (1995)
- PhD CASE studentship at PREST (1995-99)
- Panels asked to draft performance indicators (1996)
- Research Councils and Other Government Departments asked to account for implementation (1996-98)
- Royal Academy of Engineering did some case study and questionnaire work (1997)
- POST produced a review of Foresight and its impacts (1997)
- Academic work at York and Brunel universities (1997- 2000)
- OST Consultation about Lessons from First Round
- SQW contracted to develop impact indicators (1998)
- PREST/Wise Guys/SUPRA contracted to develop an evaluation framework for 2nd Foresight cycle (2000)
- Chief Scientist's Review

In the event this framework was overtaken by events as growing dissatisfaction with this cycle of the programme led the final activity in Table 2 to be instituted. The Chief Scientists Review was an internally conducted evaluation based on soliciting views from stakeholders but without any attempt to codify a systematic approach or to present detailed evidence. The conclusions however were powerful and resulted in a major change of direction (see Ian Miles' paper in this volume). There are many reasons underpinning the changes that took place but one source of explanation was that the foresight machinery in government had been constructed to optimise implementation (for example by shifting panel membership towards representatives of industry, scientific and consumer associations). When this machinery was then turned towards creating new foresight visions it was simply not equipped for the job.

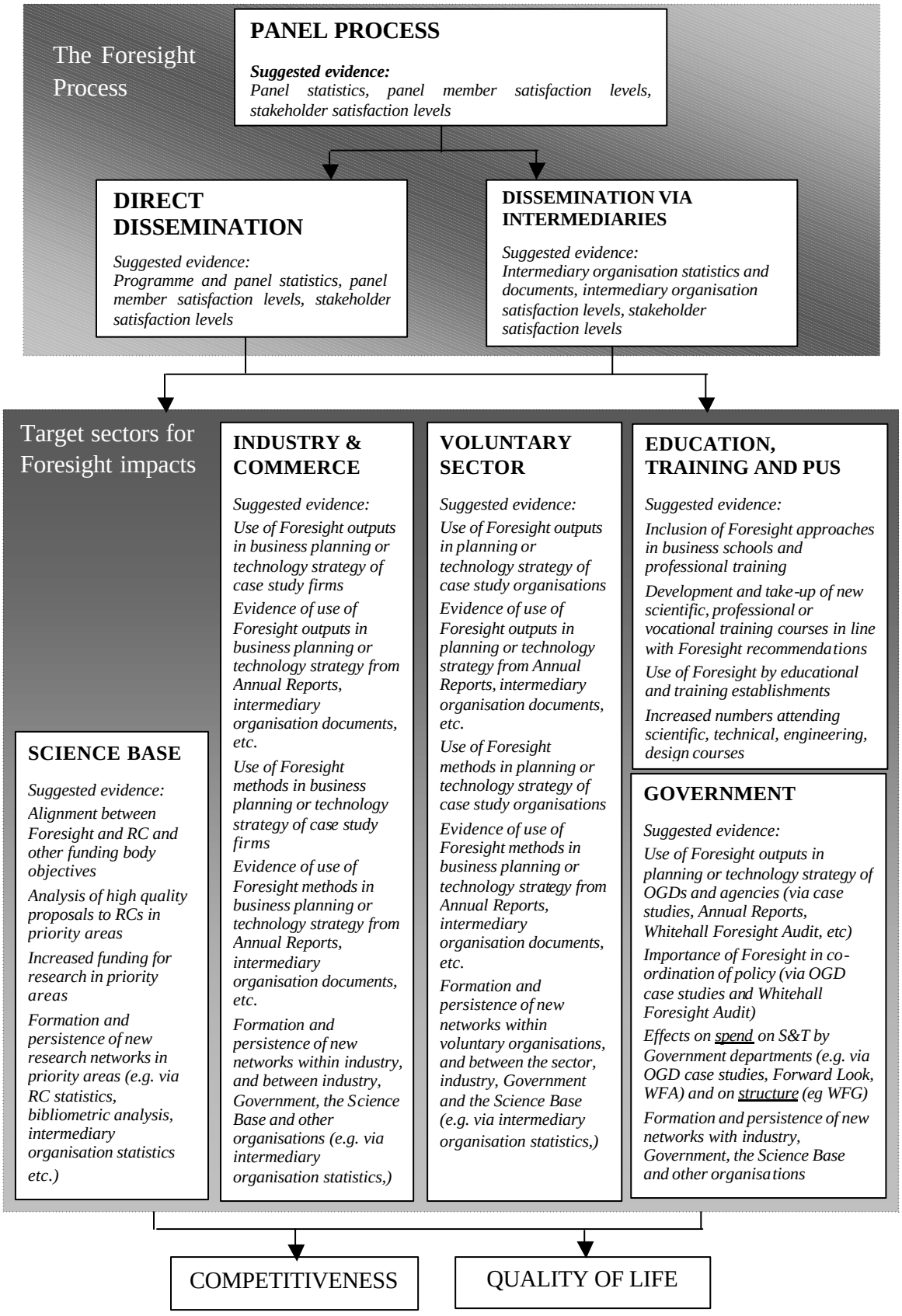
Case 2: Evaluation of German FUTUR Initiative

The second case concerns a very recent evaluation, that of the German Futur initiative (see Kerstin Cuhls' paper in this volume). This was commissioned by the responsible ministry, BMBF and was largely a process evaluation, focusing upon:

- the objectives of FUTUR, which are assumed to summarise the central assumptions upon which the exercise is based;
- the different instruments and methods with regard to their effectiveness, efficiency and interplay: and
- the process in general.

The evaluation approach was developed by ISI-Fraunhofer and involved formulating the underlying assumptions and hypotheses that underpin the ideals and conduct of Futur. These hypotheses were then 'tested' through their operationalisation into questions that could be detailed in surveys and interview protocols. Following a survey of participants a document was constructed to support an International Panel of Foresight Evaluation Experts. This panel held a one-day hearing with interviews and the Chair (the author) consulted with the Ministry as a user at the most senior level before producing the evaluation report. The limitations of this exercise were too little time and resources available and the fact that the exercise was conducted too early to pick up outcomes. However, several process-related recommendations were made and an impetus was gained for the continuation and improvement of the activity. A key finding was that the participants felt disconnected from the implementation process and to a lesser extent the programme managers responsible for implementation lacked a sense of ownership of Futur.

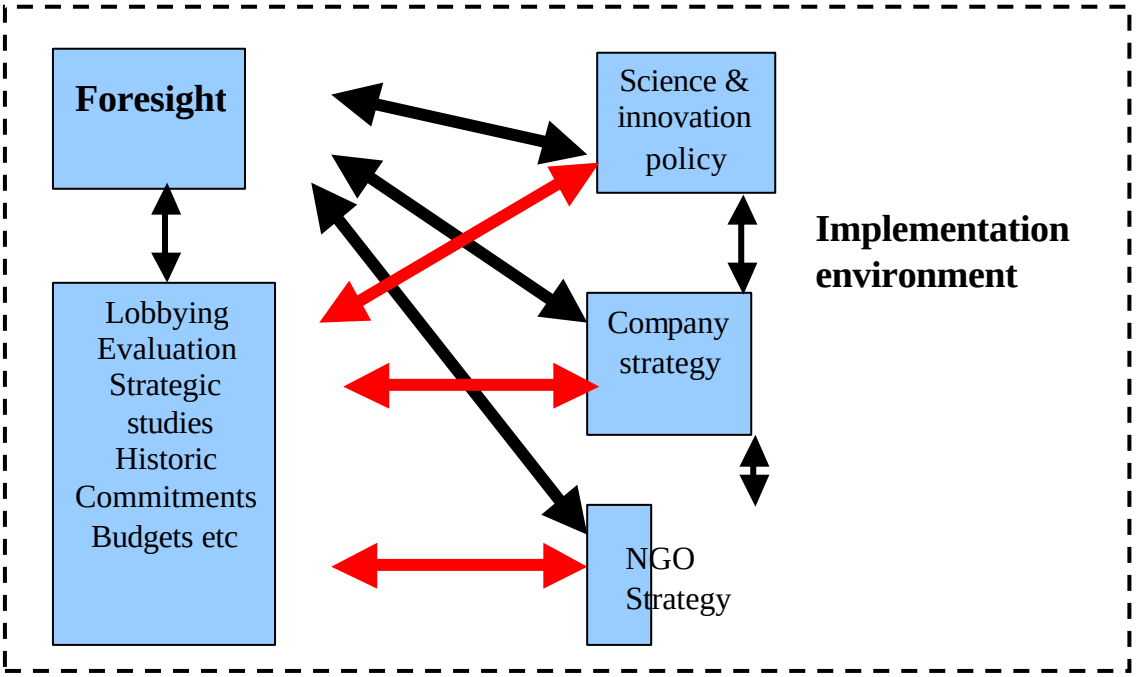
Figure 2: UK 2nd Cycle Evaluation Framework



Emerging lessons

If we return to the definition of foresight cited at the beginning of this paper and combine these with the considerations about foresight evaluation, then it can be argued that the common space and joint ownership elements in the foresight definition imply that foresight should not be in a linear relationship with implementation but rather that foresight should move into the implementation space. Figure 1 can be redrawn as Figure 3 to emphasise that foresight needs to take place inside the implementation environment, but also that foresight and implementation are interactive activities.

Figure 3: Foresight Inside the Implementation Environment



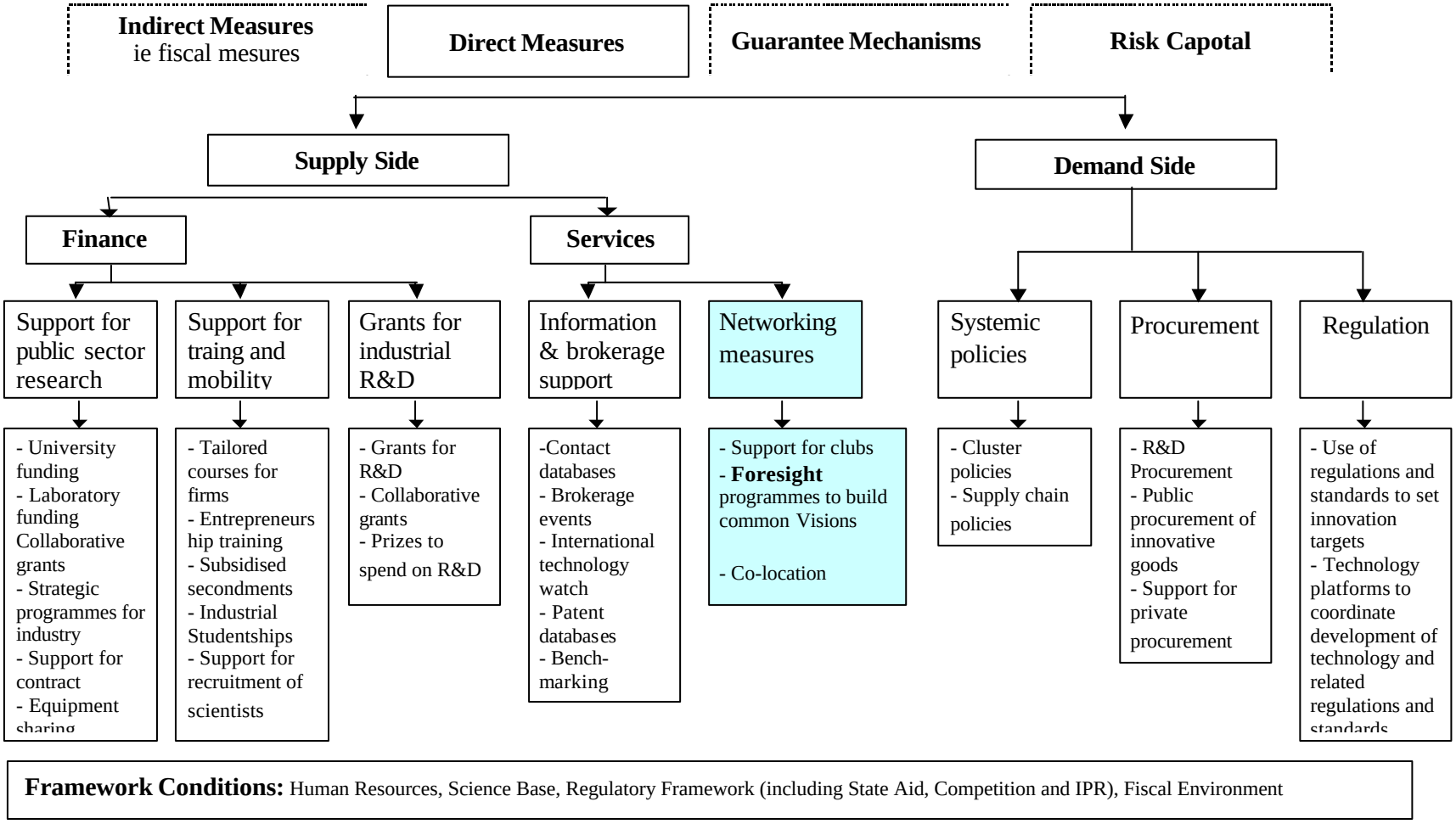
In this mature stage Foresight as an innovation policy instrument. There is a need to understand foresight in the context of a range of tools for innovation policy and hence how it can interact with, strengthen and be strengthened by combinations with other policies. Figure 4, from a current European policy exercise, illustrates how foresight sits in the array of available instruments but does not show how it can be used to enhance a range of other measures when used in combination.

To conclude and summarise:

- An integrated role for foresight in policy and strategy is needed if it is not to be marginalized and treated as an academic exercise;
- The process and implementation of foresight must both be constructed in the light of the government and company strategic processes it seeks to influence;

- For foresight to improve it must be subject to rigorous evaluation and the evaluation must feed back into new design. A key need is for evaluation to link process and effects more clearly;
- There is scope for international cooperation in evaluation in two ways
    - Sharing results to increase the speed and reduce the cost of learning; and
    - Carrying out comparative evaluation to improve the methodology and level of insight.

Figure 4: Innovation Policy Measures General Classification



(Source: Georghiou L et al, Final Report; Improving the Effectiveness of Direct Public Support Measures to Stimulate Private Investment in Research, Study Report of the Working Group on Direct Measures for Directorate General Research, European Commission, March 2002)