Evaluating Science and Society Initiatives: A Framework for Evaluation

Prepared for
Department for Innovation, Universities and Skills
Economic and Social Research Council

Joe Cullen
Fay Sullivan
Kerstin Junge

The Tavistock Institute
March 2007

Table of Contents

List of Tables

List of Figures

THE EVALUATION FRAMEWORK AND HOW TO USE IT

Why an Evaluation Framework?

This Evaluation Framework is a key output of a project commissioned jointly by the Department for Innovation, Universities and Skills (DIUS) and the Economic and Social Research Council (ESRC). The main aim of the project was to develop an approach for evaluating 'Science and Society' initiatives, and it encompassed three phases: development of a draft evaluation framework; piloting the framework in a real-world programme; and revising the Framework in the light of the pilot results. 'Science and Society' initiatives cover a broad spectrum of activities, from 'public awareness-raising' actions to educational initiatives aimed at encouraging young people to study science and related subjects in school. They reflect a number of policy agendas and actions – such as the *'Science and Innovation Investment Framework – Next Steps 2004-2014'* and the '*Science, Technology, Engineering and Mathematics (STEM) Report*' – which in turn highlight a number of issues that currently preoccupy policy-makers. These include concerns about the UK lagging behind in global research and technology (RTD) investment; a decline in the numbers of students studying Science Technology Engineering and Maths (STEM) subjects; a perception that the UK needs to improve its science, technology and engineering skills base; and the recognition that science can play a major role in contributing to improved quality of life and in promoting social inclusion.

Against this background, an evaluation framework is considered timely because:

- Although there are a large number of programmes and initiatives aimed at raising public awareness of science and supporting greater participation in STEM education and careers, there is little inter-connection and integration across these different actions. For example, the STEM Mapping Review in 2004 revealed over 470 STEM initiatives run by DfES, DTI and external agencies.
- There has been significant investment in Science and Society programmes and initiatives. The DIUS will spend £9 million in 2007 on its 'Science and Society' programme alone.
- The evidence based on the outcomes and impacts of this investment, and on 'what works', is poorly developed and varies from sector to sector.
- There is no systematic 'evaluation culture' associated with Science and Society programmes and initiatives. Evaluation practices vary in quantity and quality, and evaluation is not adequately grounded across the board in robust, rigorous concepts, models and methods. As the British Association advice to OST concluded in 2002, *"there is little published evaluation of activities and no systematic programme to assess, for example, which modes of engagement best support effective dialogue between scientists, the public and decision-makers."*

Who is it for?

The main purpose of the Framework is to help evaluators design and implement effective evaluations in this domain. The Evaluation Framework is *not* a comprehensive manual on how to conduct evaluations: there are many technical resources already available which do this.[1] Rather, it provides a common understanding of evaluation for Science and Society programmes and other initiatives. It offers a conceptual road map that can be adapted to a variety of settings. It is also aimed at policy-makers and managers of programmes and initiatives to help them specify the kinds of evaluation approaches that need to be built into programmes.

What does the Framework consist of?

The Framework combines three elements:

- A model and mapping tools to explore how political, economic, social and cultural factors shape Science and Society and, in turn, the kinds of programmes and initiatives that are implemented to promote STEM policies.
- A review and analysis of the broad spectrum of programmes and initiatives that have been implemented in the UK and elsewhere, and the kinds of approaches used to evaluate them.
- A set of generic principles, procedures and methods that can be used to implement effective evaluation.

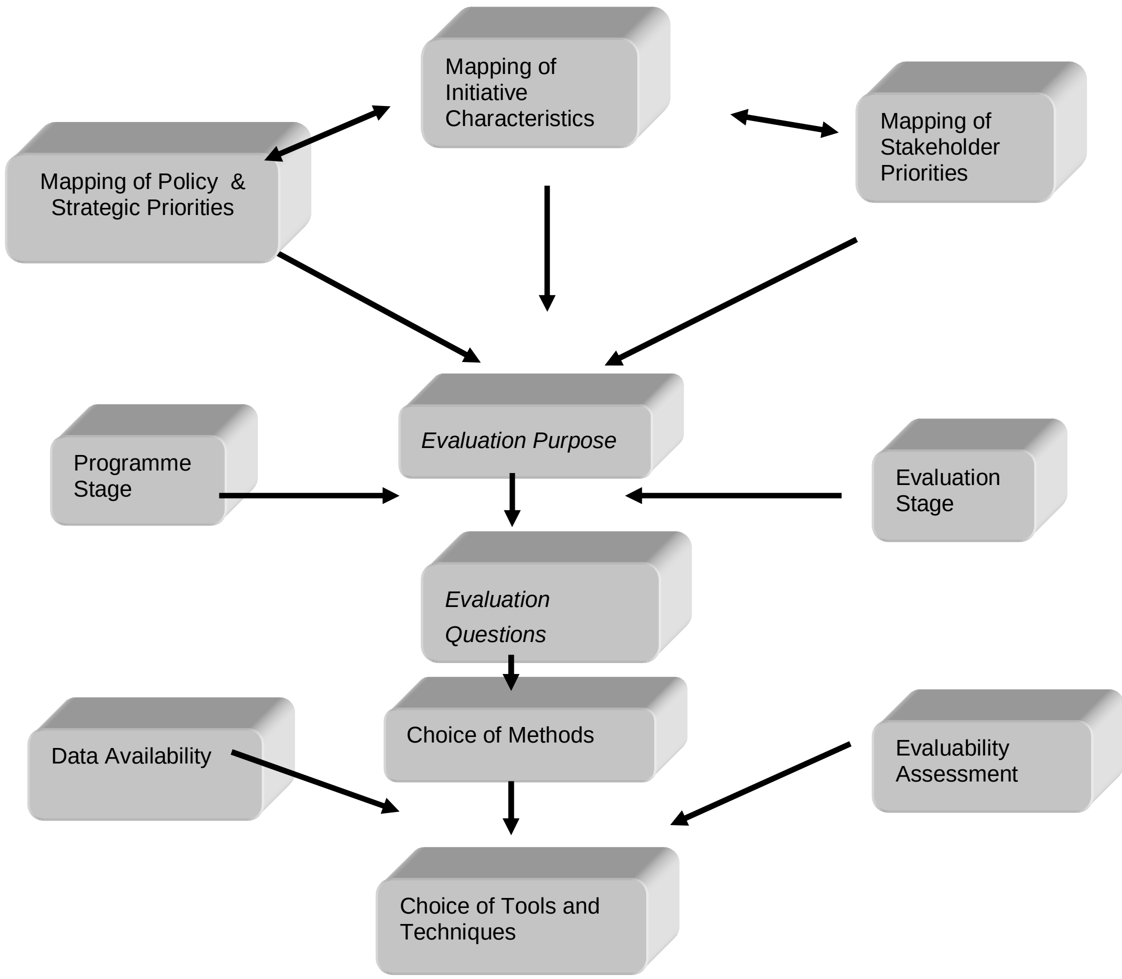
These three elements are integrated into the Framework to provide a set of procedures, guidelines, methods and examples to improve understanding of and apply key components of evaluation design in order to support effective evaluation. They help users to:

- Map the characteristics of programmes and initiatives.
- Map stakeholder priorities and needs.
- Define the object of evaluation and its purposes.
- Identify relevant evaluation questions.
- Select appropriate methods and techniques.
- Utilise evaluation results.

The figure below illustrates how these evaluation design components fit together.

[1] See for example: the Guide (formerly the MEANS Handbook) available via http://www.evalsed.info/; and RCUK's Evaluation: Practical Guidelines available via http://www.rcuk.ac.uk/news/evaluation.htm

Figure 0:1 Key Components of Evaluation Design



Adapted from The Guide (formerly The MEANS Handbook) available via http://www.evalsed.info

Where can I find the relevant information?

The Framework is split into three parts, corresponding to the three elements described above:

- Part 1 examines how political, economic, social and cultural factors shape Science and Society and, in turn, the kinds of programmes and initiatives that are implemented to promote STEM policies.
- Part 2 provides a review of the characteristics of typical Science and Society programmes and initiatives, and the kinds of approaches used to evaluate them.
- Part 3 provides principles, procedures, methods and examples to support evaluation design.

The Framework in total consists of eight sections and three annexes. These sections broadly correspond to the components and phases of evaluation design shown in the illustration above. The table below summarises the contents of each section, what it covers and how it is intended to be used.

Table 0-1 Summary of Sections

PART 1: Understanding the domain			
Section	**Title**	**What this section covers**	**You should read this if…**
1	Introduction	The purposes, scope and objectives of the Framework.	You want to know the background to this study and what the Framework is for.
2	Developing an Evaluation Framework for Science and Society Initiatives	Develops a model for understanding the 'object' of evaluation. Based on the interaction between 'proximal' and 'distal' dynamics.	You are interested in the evaluation theory underlying the Framework.
3	Understanding Science and Society: the Domain	Explores how science is 'culturally constructed'. Provides a mapping tool to help evaluators understand the 'cultural logic' of an initiative.	You are interested in the social and cultural processes that affect how and why Science and Society policies and initiatives are developed and implemented.
4	Understanding Science and Society: the Initiatives	Reviews the literature on Science and Society approaches. Provides a typology of initiatives. Reviews how they are evaluated.	You want to situate a particular initiative within the spectrum of current programmes and initiatives.
PART 2: Building an Evaluation Framework			
Section	**Title**	**What this section covers**	**You should read this if…**
5	The Evaluation Framework: Building Blocks	Provides a template for evaluating Science and Society initiatives.	You want to find out about the evaluation procedures required to design an evaluation.
PART 3: Applying the Framework			
Section	**Title**	**What this section covers**	**You should read this if…..**
6	Making the Evaluation Framework Work in Practice	Adapts the evaluation building blocks in Part 2 to the characteristics of the Science and Society domain.	You want to design and carry out an effective evaluation of a Science and Society initiative.
7	Typical Science and Society Evaluation Scenarios	Provides examples of a range of different Science and Society initiatives and how to evaluate them.	You require 'benchmarks' and good practices to help you choose the appropriate evaluation approach

			and methods.
8	Illustration: Evaluation approach for 'Sciencewise'	Provides a detailed evaluation design for a real world initiative.	You want a step-by-step illustration of how to design an evaluation in practice.
ANNEXES			
Section	**Title**	**What this section covers**	**You should read this if…**
Annex I	Innovative Methodology	Explores the application of two particular approaches to evaluation: ex-ante and experimental approaches.	You are interested in how to carry out 'prospective' evaluations, longitudinal evaluations and randomised controlled trials.
Annex II	Selected References	Provides key references on evaluation.	You want further details of the evaluation theories and practices that have informed the Framework.
Annex III	Glossary	Provides definitions of key evaluation terms.	You want further explanation of the evaluation concepts and terms used in the Framework.

Note on Annex I

The study included work on 'innovative methodologies' that could be applied within the Science and Society domain. We use the term 'innovative' here not in the sense that there is anything new about the methods discussed, rather that they have been relatively under-utilised – and arguably sometimes miss-applied – in the Science and Society field. The methods described in Annex I – ex-ante evaluation, longitudinal evaluation and randomised controlled trials - have been chosen in collaboration with the commissioners of this project, as it is felt that they offer particular promise for addressing questions of interest to funders of Science and Society initiatives. On the basis of our review of evaluation approaches used in the domain, there is some evidence to suggest that ex-ante evaluation has been relatively neglected by evaluators, and many experts take the view that incorporating ex-ante evaluation into programme and initiative design could considerably reduce problems and issues that are subsequently encountered as the programme or initiative is implemented. Similarly, many experts agree that 'experimental' approaches, including longitudinal evaluations and randomised controlled trials, offer the most robust and 'scientific' ways of assessing the longer term impacts of programmes and initiatives, and of demonstrating 'cause and effect'. Yet, they are very difficult to design and apply, and are often used in the wrong circumstances and for the wrong purposes. Against this background, Annex I outlines the main features of these approaches, what they entail and under what conditions they should be used.

1. INTRODUCTION

1.1. The project to design an evaluation framework

In February 2006, the Tavistock Institute was commissioned by the then Office of Science and Technology (now the Department for Innovation, Universities and Skills, DIUS), in association with the Economic and Social Research Council (ESRC), to design, pilot and then further refine an evaluation framework for use in evaluating Science and Society initiatives.

A first version of the Evaluation Framework was drawn up by the Tavistock Institute on the basis of: a literature review of current evaluation practice in the Science and Society field; interviews with evaluation experts and evaluators and commissioners of Science and Society initiatives. This version of the Evaluation Framework was then pilot tested in a real-world evaluation, thanks to SETNET (the Science, Engineering, Technology, and Mathematics Network), who were generous enough to let us use them as our guinea pigs.[2] Learning from the application of the Evaluation Framework in this setting has resulted in significant revision of the original version, the results of which are presented in this report.

1.2. What is the Evaluation Framework?

The Evaluation Framework is *not* a comprehensive manual on how to conduct evaluations: there are many technical resources already available which do this.[3] Rather, it provides a common understanding of evaluation for Science and Society programmes and other initiatives. It offers a conceptual road map that can be adapted to a variety of settings.

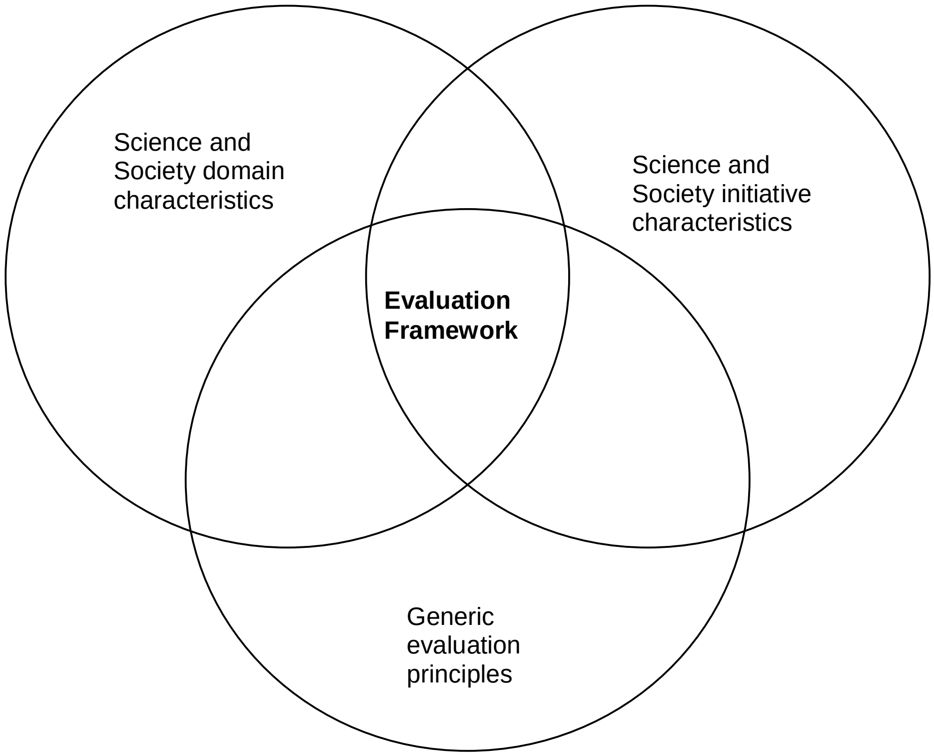
There are three essential ingredients that make up the Evaluation Framework, see Figure 1:1:

- An understanding of the nature of the Science and Society domain.
- An understanding of the nature of specific Science and Society initiatives.
- An understanding of best practice in evaluation.

These elements are combined to produce an Evaluation Framework relevant to evaluations in the Science and Society field.

[2] A report of the pilot evaluation, Cullen J, Sullivan F, and Junge J *Evaluating Science in Society Initiatives: SETNET Evaluation Framework Pilot Report*, will shortly be published by the OSI on its website.
[3] See for example: the Guide (formerly the MEANS Handbook) available via http://www.evalsed.info/; and RCUK's Evaluation: Practical Guidelines available via http://www.rcuk.ac.uk/news/evaluation.htm

Figure 1:1 The three essential ingredients of the Evaluation Framework



1.3. Using this report

The report is divided into three parts. The first part both describes the development of various elements of the Evaluation Framework and begins the work of creating the understanding of the Science and Society domain and Science and Society initiatives that is necessary for successful evaluation in this field. The second part of the report brings this understanding together with best practice in evaluation to produce the Evaluation Framework itself. Part three provides some practical tips and examples of how the Evaluation Framework can be applied in practice.

In Annex I we present a discussion of ex-ante evaluation and longitudinal or prospective research methods, areas which both we and the commissioners of this project felt were potentially highly relevant to the evaluation of Science and Society initiatives but have not yet been explored in the field in great detail. Annex II provides a set of key references for further study, and Annex III provides a glossary of the terms used in the Framework.

PART I: UNDERSTANDING THE DOMAIN

2. DEVELOPING AN EVALUATION FRAMEWORK FOR SCIENCE AND SOCIETY INITIATIVES

2.1. What this section is about

This section sets out a conceptual model that depicts the interactions between the structures and processes that shape science at the broader societal level and those structures and processes that shape particular Science and Society initiatives on the ground. It is intended to help evaluators gain a better understanding of how 'external' factors – for example, popular culture – help to inform people's perceptions of science and scientific knowledge. This background information can be used to inform the kinds of evaluation questions that need to be asked in understanding the purposes, objectives and intended outcomes of a Science and Society programme or initiative.

2.2. Model of the Science and Society environment

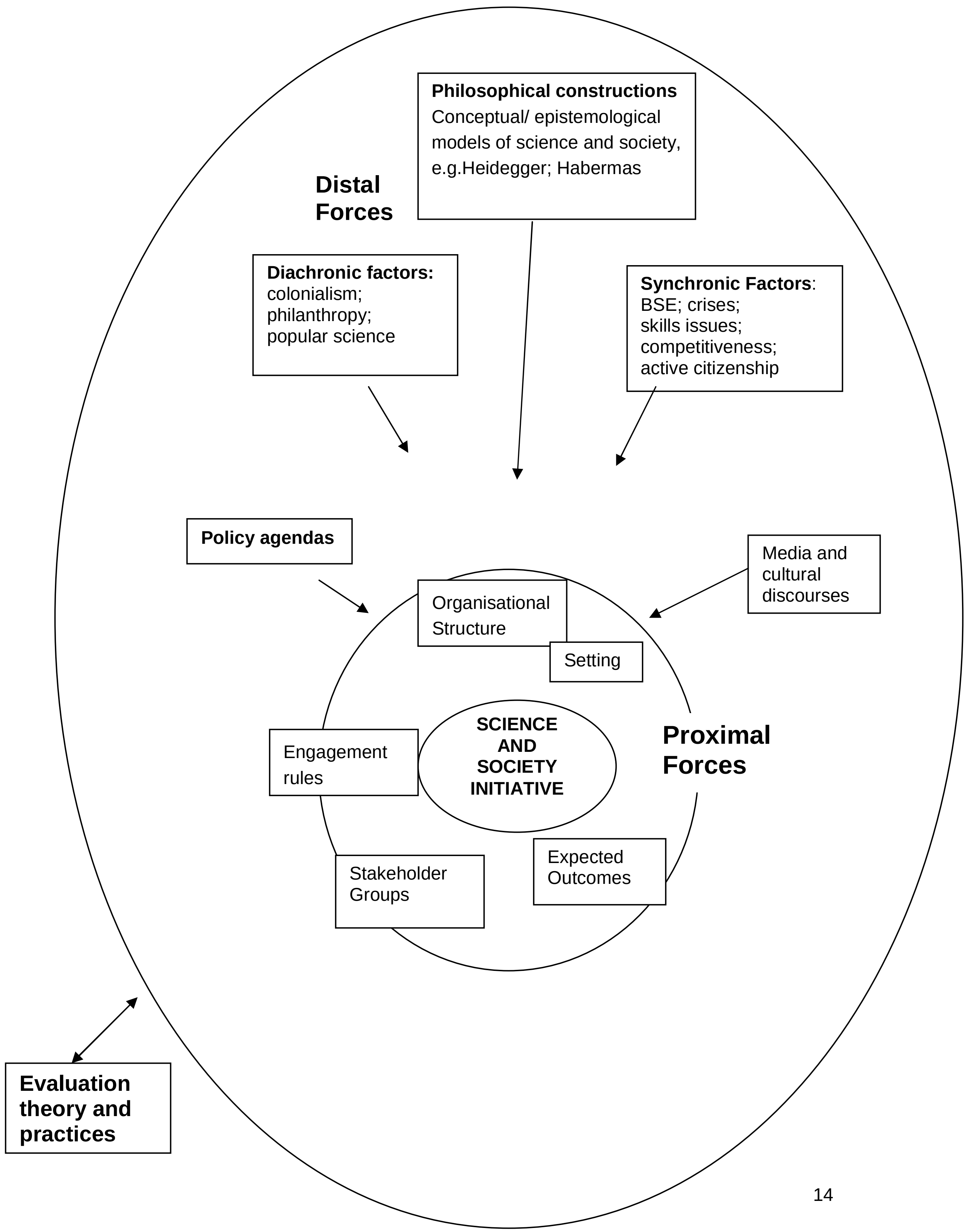
In the following sections of this report we develop an Evaluation Framework and toolkit for Science and Society initiatives based on the integration of the three elements described in the introduction. In essence, the approach used to develop the Framework entails applying current evaluation theory and practice to a conceptual framework that models the interactions between the structures and processes that shape science at the broader societal level and those structures and processes that shape particular Science and Society initiatives on the ground. The model used is shown in Figure 1.

Figure 1 suggests that the 'object of evaluation' (that is, the policy, programme, project or other action to be evaluated) at the centre of the diagram needs to be defined by applying evaluation theory and practice to an understanding of the relationships between the 'distal' and 'proximal' forces that govern how science is constructed within its social context.

Distal forces can be defined as the diachronic and synchronic (historical and societal) factors that shape constructions of Science and Society at the macro level. They would include different philosophical, conceptual and epistemological perspectives on what science is and what role it plays in society. They would reflect broad cultural 'patrimonies' at the national level, i.e. the ways in which a nation has developed its own distinctive 'science culture' over time. For example, in the case of the UK, some aspects of public perspectives of science reflect the legacy and influence that Victorian engineers such as Brunel had on the public imagination. Distal forces also include policies aimed at promoting science and how these link to other policy agendas, such as economic competitiveness.

Proximal forces are defined as the factors that directly shape the immediate implementation of a Science and Society activity within its particular context, i.e. within a local setting or scenario. They include physical elements (such as the space in which the action takes place); normative elements (such as the rules and procedures governing the activity); socio-cultural elements (such as partnerships and networks); and psycho-social elements (such as whether the activity is appropriate, given the profiles of the participants involved).

Figure 2:1 Model of structures and processes shaping Science and Society initiatives and the environment in which they operate



3. UNDERSTANDING SCIENCE AND SOCIETY – THE DOMAIN

3.1. What this section is about

This section presents the results of a literature review of the concepts, theories and practices in the Science and Society domain. It looks at how opinions and perceptions of science are shaped and at the different kinds of models that have been used to design and deliver Science and Society programmes and initiatives. Drawing on the results of the review, this section identifies seven key factors that need to be considered in evaluation design. It presents a 'mapping tool' to help evaluators understand and describe how a programme or initiative reflects particular 'social and cultural constructions' of Science and Society, and how this relates to evaluation design.

3.2. Introduction

In this section we look at those factors and processes at the broader macro level that shape how science is constructed within its societal context and what this means for evaluation design. First we present a brief review of the main concepts and issues, drawn from a review of the literature. The second part of this section provides a 'mapping tool' to help provide inputs to subsequent evaluation design.

3.3. Concepts and issues: the social construction of science

If people behaved like electrons (or at least in ways similar to what is commonly assumed to be the general behaviour of electrons) then there would be no need for Science and Society programmes. Like the behaviour of electrons, people's behaviour with regard to science could be predicted. Everyone would share the same beliefs about its benefits and value. Science Week would be unnecessary, since public trust and confidence in science would remain undiluted and constant. However, this functionalist perspective on human behaviour, like the logical positivist perspective on scientific behaviour, is at odds with reality. The reality is that science does not exhibit one single, cohesive form that is recognised by everybody – it exhibits many different forms. This is because science is socially and culturally constructed. This does not simply mean that science is viewed differently by different people at different times. The social and cultural construction of science implies rather that the fundamental nature of science itself – its very essence – will be shaped by prevailing synchronic and diachronic processes. In other words, it isn't just that the meaning attached to science in a Liverpool council estate in the 1980s is different from the meaning attached to science at a Royal Society lecture in London in 2006. It is that the science itself is different.

Our starting position for developing a framework, methodology and tools to evaluate Science and Society programmes, initiatives and actions is, therefore, to recognise, understand and classify how science is socially and culturally constructed. This is no easy task, since the construction of science is part of a broader, complex set of processes through which knowledge is created, disseminated and utilised through what are termed 'communicative practices'. At one level, the construction of science is shaped through philosophical and sociological discourses. At one end of the spectrum of Science and Society discourses, dystopian post-modernists like Heidegger and Habermas saw science as part of the broader, all-pervading process of 'technicisation' – inexorable, overwhelming and destructive. In this 'essentialist' perspective, science is constructed as an 'instrumental' force. As Feenberg (1996) observes, Heidegger's view is that technology is relentlessly overtaking us. Modern society is engaged in the transformation of the entire world, ourselves included, into 'standing reserves', raw materials mobilised in technical processes. Scientific knowledge and practices are the fuel that runs technical processes. Moreover, science and technology present what Feenberg (1996) calls 'civilising choices'. A current example is the debate over sustainable energy technologies. The solutions that science provides to address climate change problems and sustainable energy demands will inevitably reflect choices. Some theorists argue that the main choices are between supporting the continuing supremacy and control of the developed over the developing world or contributing to radical changes in global economics. For example, it has been argued that the 'civilising choices' scientists make in relation to sustainable technologies are currently being shaped by policy discourses aimed at supporting the UK and its EU partners in their drive to become the 'most competitive economy in the world', in line with the objectives of the Lisbon Conference.[4]

At the other end of the spectrum, in successive elaborations of the concept of 'dialogic reflexivity', Anthony Giddens has argued that the increasing pervasiveness of the 'Knowledge Society' has opened up new opportunities to enable ordinary people to re-colonise ground that for so long has remained exclusively the terrain of 'experts'. Giddens suggests that the Knowledge Society has allowed ordinary people to have a say in the creation of new knowledge, using the Internet and related technologies such as weblogs (Giddens, 1991; 1994; 2000). Individuals can become their own personal 'social laboratories', carrying out everyday 'experiments with the self', and contributing to the creation and application of new scientific knowledge and processes, including those that will have 'positive' or beneficial effects for civilisation.

It is difficult to judge how and in what ways these philosophical (or 'intellectual') discourses affect public constructions of science. What is clear is that 'public' constructions of science have distinctive diachronic and synchronic variability – that is, they reflect both historical processes and cultural dynamics in particular moments in time. So there is never one single homogenised public perception of what science is and does. It has to be acknowledged that science communication activities have always recognised that there is more than one public. A recent review conducted jointly by the then OST and the Wellcome Trust, for example, assessed the effectiveness of a range of Science and Society initiatives with regard to how they approached the different demographic and socio-economic features of different audiences. The research showed that, although perceptions of science are related to variables such as gender, age and economic status, "basic personal attitudes are the primary factor that determines an individual's attitude to science and engineering".[5] The OST/Wellcome research identified six distinctive 'clusters' of attitudes to science

---

[4] European Council, Lisbon, March 2000.

[5] Office of Science and Technology and the Wellcome Trust, Science and the Public: A Review of Science Communication and Public Attitudes to Science in Britain (London, October 2000).

and argued that programmes such as science awareness-raising initiatives could be more successful if they incorporated 'hooks' targeted at each cluster's interests. However, what the OST/Wellcome research fails to address are the complex ways in which discourses around science reflect more fundamental structural features of society – particularly power structures and their historical development. As Miller (2001) puts it:

> *"Public attitudes to science, at least insofar as they are reflected in the media, showed periods of great adulation and expectation immediately after the war, followed by disappointment and even hostility, giving way to a generally ambiguous viewpoint. Alongside these 'mood swings' there was a tendency for scientists to retreat into their shells, frowning on those who ventured onto the public stage, thus mirroring attitudes of their counterparts in the United States."* [6]

The ebb and flow of society's love affair with science highlights the contradictory and ambivalent nature of how science is publicly constructed. Because of this ambivalence and paradox, political and policy processes become extremely important factors in the construction of science. In recent times in the UK, political and policy discourses around science have largely been dictated by crises such as BSE and by a more general increasing public distrust of science and scientists. This has been associated with rapid advances in scientific knowledge and a corresponding low level of information available to the public about the nature of these new advances. As the House of Lords Select Committee on Science and Technology put it at the beginning of its 2000 report on Science and Society:

> *"Society's relationship with science is in a critical phase. Science today is exciting, and full of opportunities. Yet public confidence in scientific advice to Government has been rocked by BSE; and many people are uneasy about the rapid advance of areas such as biotechnology and IT – even though for everyday purposes they take science and technology for granted. This crisis of confidence is of great importance both to British society and to British science."*

The emphasis of the Lords' report was on increased openness and transparency in the treatment of scientific advice, the recognition of scientific uncertainty and the legitimacy of public values and concerns. What is striking about the recent history of Science and Society policy, however, has been the over-riding assumption that public concerns were simply a matter of providing the correct information. As Irwin (2001) points out, Lord Sainsbury perfectly represents the assumption that better communication will resolve problems of public confidence in science in his proposal of a solution to anxieties over BSE and similar 'crises of science':

> *"To remedy this Government has to ensure that not only are its systems appropriate, but that their existence and role are communicated. To restore public confidence in the Government's use*

[6] Miller, S (2001) Public understanding of science at the crossroads, Public Understand. Sci. 10 (2001) 115–120

> *of scientific advice required people to understand the mechanisms used to arrive at decisions and accept that those were appropriate and based on sound principles."*

This position highlights the extent to which the so-called 'deficit model' has come to dominate policy and practice around Science and Society programmes. As Miller (2001) acknowledges, the deficit model assumed "public deficiency, but scientific sufficiency". It adopted a one-way, top-down communication process, in which scientists (with all the required information) filled the knowledge vacuum in the scientifically illiterate general public as they saw fit. The model portrays a simple causal chain between information provision and societal change: more information means better scientifically-informed people which means more responsible decision-making; a richer and more marketable skills set; more active social and political engagement; and a better society. Yet this perspective is highly contested (Facer et al, 2001; Supple, 1999) and the notion that simple exposure to information broadens horizons and raises understanding is not supported by the evidence (Buckingham, 1999).

The real point about the 'deficit model', as Irwin observes, is not how it constructs science within society, but, more importantly, how politics and policy construct the 'scientific citizen' within prevailing policy and decision processes. He suggests that there are three fundamental frameworks that govern the relationship between science and citizenship. These can be understood as competing "technologies of community". The first one can be conveniently labelled a *social research* framing of science-citizen relations. This has sought to achieve both depth and representativeness across the population and to encapsulate public views around issues that are closely tied to prevailing institutional agenda and working practices. The second framework can be termed the *deliberative democracy* model of direct discussion and engagement (represented, for example, by the *Citizen Foresight* project). This model is shaped at least partially by (selected) citizens themselves, and is linked to established active citizenship models such as citizens' juries and consensus conferences. A third, and more embryonic, framework can be viewed as a *qualitative and localised* model that seeks to consider public assessments as expressions of 'discursive' and 'communicative practices'.

Irwin's analysis re-emphasises the crucial power dimension in the construction of science within society. Put simply, the argument is that the 'dominant' construction of Science and Society will be determined by the outcome of a 'contest of meanings' between competing power constituencies. Elzinga and Jameson (1995) contend that four main cultural groups – academic, industrial, bureaucratic and civic – can be identified as the key protagonists in the struggle for control over scientific discourses. Each group competes for resources and influence, and seeks to steer science and technology in particular directions, based on its particular doctrinal assumptions, ideological position, ideals of science and desired pragmatic outcomes, as shown in Table 3-1.

Table 3-1 The cultures of science policy

Culture	Mode of public assessment	Core concern	Pragmatic implication
Academic	Appreciation	Cultural achievement	Scientific autonomy
Industrial	Acceptance	Technological opportunity	Competitive advantage
Bureaucratic	Trust	Epistemological authority	Effective political action
Civic	Responsibility	Social consequences	Accountability

Excerpt from Elzinga and Jamison, 1995

The academic culture will place highest value on public appreciation and prestige of science. The industrial culture will primarily be concerned that science and technology be perceived as unequivocally progressive forces, facilitating a wholesale acceptance in advance of any technological innovation which may justify the unrestricted use of science and technology as sources of competitive advantage. The bureaucratic culture may be primarily concerned with science as a source of certified knowledge, by which options for effective political action may be developed and justified. Lastly, the civic culture will be concerned that the institutions and practices of science exhibit a strong sense of social responsibility and responsiveness.

Of course, in practice, there is considerable cross-cultural blending in the construction of scientific knowledge. For example, academic cultures are never 'pure', and their agendas are increasingly shaped and implemented within the context of partnerships that include industrial, bureaucratic and civic cultures. It is more realistic to envisage scientific discourses as the product of a continual interplay between loosely-defined partnerships and alliances that interact with both relatively stable hegemonic power structures and more volatile, short term dynamics that may reflect localised factors. It might be more useful, therefore, to think about the construction of Science and Society more in terms of 'scenarios' rather than particular stakeholder groups and power structures. One kind of scenario-based framework is adopted by Martin and Richards (1995) in their contention that constructions of science can be more effectively understood in terms of how controversies evolve and are managed. This acknowledges that the traditional image of the scientist as rational, disinterested guardian of truth and public interest has long been abandoned, along with the decline in public trust in the infallibility of science and an increasing demand for greater involvement by laypeople in scientific decision-making. Martin and Richards identify four broad approaches used to analyse how science is constructed around public involvement in the management and resolution of controversy. Firstly, the 'positivist' approach adopts the position that controversy is the result of error on the part of those people who divert form the scientific orthodoxy. The researcher's or evaluator's task is, therefore, to establish why this divergence has occurred and how it can be resolved. In the second 'group politics' approach, the research or evaluation methodology focuses on analysing the interactions of the plurality of groups (government agencies, citizens' panels, professional bodies and so on) whose actions give rise to the scientific dispute or controversy in question. In contrast, the constructivist approach concentrates on how scientific knowledge is developed through the different constructions of 'truth' around scientific issues. Finally, the social structural approach considers the social context of science in terms of relationships between groups and structures that constitute social space, in particular those structures, like economic and social class, that most directly mediate power relations in society.

Research on scientific controversies would appear to reinforce the picture of a science that is socially constructed through contextualisation. In its broadest sense, there is plenty of evidence to support the view that science is constructed according to cultural 'patrimonies' that are broadly expressed at the national level. For example, in contrast to the UK, the evidence of four decades of research into Science and Society programmes suggests in the US that there is a strong and continuing public belief in the value of scientific research for economic prosperity and for the quality of life. Even though there are some continuing reservations about the pace of change engendered by science and technology and the relationship between science and faith, the public consistently reconciles these differing perceptions in favour of science (Millar, 2004). Similarly, regular monitoring of public attitudes towards science and technology carried out by the Japanese Ministry (NISTEP) shows a high level of general interest in science and technology and, as in the US case, a relatively constant endorsement of the positive value of science, although increasing specialisation led to perceptions that science and scientists have become more and more closed off to the public in recent years.[7] In broad terms, the evolution of particular 'science cultures' on national lines is reflected in, and can be measured by, indicators such as relative scores on 'scientific literacy' scales, used, for example, in successive Eurobarometer and OECD PISA surveys over the past decade.[8]

However, there is strong evidence that local 'context' is just as, if not more important when it relates science to particular situations and instances that are embedded in everyday life. Millar (2001), for example, argues that a lot of the science the public engages with is in the 'controversy' mode, typically involving acute, and potentially threatening, situations. This type of science, he suggests, is of a "science-in-the-making" variety that is still being "socialised" by the scientific community. "Textbook" scientific certainties rarely hit the headlines to grab the public's attention. These considerations support the idea of a 'contextual approach' to the public understanding of science. Here, the creation of scientific knowledge and of understandings of science are collaborative and negotiated dialogues, in which scientists have scientific facts at their disposal, but the members of the public concerned have local knowledge and an understanding of, and personal interest in, the problems to be solved. Practical experiments in this approach include consensus conferences, citizens' juries and 'science shops'.

To summarise, a review of the literature suggests that an effective evaluation approach needs to take account of the following considerations:

- There is no single consensus about the role of Science and Society. Science is constantly being constructed and reconstituted as society itself evolves.

[7] Mitusishi S et al (2001) A new way to communicate science to the public: the creation of the Science Library, ubb. Und. Sci., 10, 2
[8] see OECD Pisa Scientific Literacy Assessment Framework, 2006:; Eurobarometer

- Conceptions about science and attitudes towards how science can contribute to personal and social well-being vary in complex ways.
- This variation is shaped at the macro level through broad patterns and processes that create fundamental social structures themselves, for example, economic and social status.
- It is also shaped by historical processes, including key historical ‘events’. People’s attitudes towards and expectations of science are also reflected in terms of individual personality profiles.
- To some extent, constructions of science reflect the power relations and interactions between broad stakeholder alliances, for example, representing the ‘academic’, industrial, civic and bureaucratic constituencies.
- However, these power structures are far from homogenised; they are fluid and constantly evolving, and present quite complex inter-relations and partnerships across constituencies. An example of this is the spectrum of conceptual positions on Science and Society represented by philosophers of science, sociologists and other intellectuals.
- The social construction of science is also framed through contextualisation. This involves complex dialogue and negotiation between representatives of the formal scientific knowledge base and ‘lay people’ who have local knowledge and concerns. However, the process of contextualisation increasingly reflects new forms of social relationships and dialogue promoted through the increasing utilisation of ‘knowledge society technologies’.

We have argued that there is no one, timeless construction of science. We have, therefore, not offered a definitive view of what this construction looks like. Rather, we have presented a number of factors which influence its construction, and draw evaluators’ attention to the nature of the Science and Society domain in order that they can better evaluate Science and Society initiatives.

3.4. Mapping tool: Distal level

There are a number of reasons why it is useful for evaluators to understand how and in what ways science reflects different perspectives and constructions within society. To take one example, the BSE crisis showed that politicians and policy-makers can not only misunderstand the fundamental nature and focus of public anxieties about science at a particular time and place, they can also fail to understand how these anxieties are specifically expressed, and, therefore, they can fail to adequately diagnose the appropriate actions necessary. Against this background, evaluators need to understand the nature of a particular construction of science, so that they can then assess the appropriateness and ‘goodness of fit’ between this construction and the choices made in an intervention.

There are a number of factors about how a Science and Society intervention is ‘constructed’ that are of particular importance in evaluation terms. These include the ‘cultural logic’ or ‘vision’ that lies at the heart of an intervention; the ‘theory of change’ represented by the intervention; the extent to which the ‘vision’ of the intervention supports or opposes prevailing policy agendas; and the epistemological and philosophical perspective on science it represents.

'Cultural logic' can be defined as four key elements which reflect the 'vision' of an initiative in relation to:

- What it sets out to do in terms of the aims and objectives ascribed to it by key stakeholders (universalisation).
- How coherent the vision is, in terms of the extent to which it is shared by stakeholders, (closure).
- The practical choices made to realise the vision and its objectives (specification).
- The emphasis on and capacity for learning from the innovation or initiative, and hence its capacity to evolve and adapt in response to external and internal influences (situational change).

The cultural logic of a Science and Society initiative may reflect, for example, particular notions of national economic and social objectives and the role that science plays in these. These may emphasise: human capital constructions, about driving the economy forward through science; social capital constructions, focusing on the intrinsic value of science and its contribution to promoting individual and social development; and humanist (needs-focused) constructions, focusing on human and social problems, and the tools science can offer to solve them.

In turn, the cultural logic of a Science and Society initiative will be underpinned by a 'theory of change' or mechanism via which the initiative is expected to effect its desired impact. The changes envisaged and incorporated into the 'logic' of a programme or initiative can vary significantly. They can be pitched at a broad generic level, for example, improving citizens' awareness of science so they can have a more informed view on potential decisions, or at a highly instrumental level, for example, increasing the proportion of black and ethnic minority community representatives in the physics teaching profession. Theories of change will reflect broader level historical antecedents that contextualise the initiative within the events that led to a particular set of policies governing Science and Society. The theory of change links the overall purposes of the initiative and its intended outcomes and thus acts as a bridge between the distal and proximal dimensions.

In addition to enabling an assessment of the 'goodness of fit' between the construction of science embodied by the initiative and that of the environment within which the initiative will function, understanding how Science and Society interventions are constructed in terms of these elements will help the evaluation to specify, for example:

- The 'object' of the evaluation (what is the 'unit of analysis' to be evaluated).
- The purposes and scope of the evaluation (for example, human capital; social capital focused).
- The range of stakeholders that need to be involved.
- The type of evidence that will be accepted.
- The questions that need to be asked and the criteria that need to be used to draw conclusions.

Table 3-2 provides a mapping tool to help evaluators understand and articulate how an intervention is constructed and how this relates to evaluation design.

Table 3-2 Mapping tool – distal level

Dimension	Data required	Inputs to evaluation
Cultural logic (universalisation)	Overall vision and purpose of initiative Goals and aims	Identifying the purposes of the evaluation, e.g. descriptive; analytical; operational Specification of evaluation questions and criteria, e.g. Cultural relevance – in what ways does the initiative support or run counter to how this aspect of science has been viewed over time?
(closure)	Stakeholder representation and power relationships Degree of consensus / contention of subject matter Level of current knowledge	Identifying the purposes of the evaluation, e.g. developmental; consensus-building Specifying the evaluation mode, e.g. external / objective; self-evaluation; peer support Specification of evaluation questions and criteria, e.g. Accessibility – does the intervention exclude particular groups?
(specification)	Specification – mapping the methodological and operational choices made	Specification of evaluation questions and criteria, e.g. User-friendliness and usability – are users satisfied with the vehicle chosen to deliver the intervention?
(situational change)	How learning and transferability of knowledge is reflected	Identifying the purposes of the evaluation, e.g. developmental; consensus-building
Theory of change	What model of change underpins the vision of the intervention and on what level is change anticipated (individual; societal)?	Identifying the purposes of the evaluation, e.g. descriptive; analytical; operational Specification of evaluation questions, criteria and measurements, e.g. to what degree has the intervention increased average scores on science literacy rating scales?
Initiating circumstances	Factors influencing origination of the intervention	Identifying the purposes of the evaluation, e.g. descriptive; analytical; operational
Policy agenda and focus	Link to specific science policies Relationship and integration with other policy areas	Specification of evaluation questions and criteria e.g. in what ways has the initiative supported science policy? Specification of evaluation questions and criteria, e.g. what value has this initiative contributed to employment policy?
Philosophical and epistemological focus	The kinds of 'evidence' that are considered credible	Helping to decide on the kinds of methods and techniques used to gather data

4. UNDERSTANDING SCIENCE AND SOCIETY – THE INITIATIVES

4.1. What this section is about

The first part of this section presents the results of a review and analysis of a range of typical Science and Society programmes and initiatives that have been carried out in the UK and further afield. On the basis of this review, a typology of programmes and initiatives is developed, which outlines the key characteristics of the different types identified. The second part presents a review of the kinds of approaches that have been used to evaluate these different types of programmes and initiatives. Drawing together the results of the two reviews, the section concludes with a 'mapping tool' to help evaluators match evaluation design to the characteristics of particular programmes and initiatives.

4.2. Introduction

The preceding section focused on understanding and mapping how Science and Society is constructed from the broader social environment (distal level). It looked in particular at the 'cultural logic' of Science and Society initiatives from this broad 'external' perspective. In contrast, this section starts from the other end of the spectrum, the 'proximal level', and concentrates on the ***internal*** logic of a specific initiative. In this part of the study we were interested, firstly, in finding out whether distinctive types or clusters of initiatives could be identified and distinguished on the basis of their characteristic features and, secondly, whether these different types of initiative were associated with distinctive modes and methods of evaluation. The first part of this section, therefore, presents an analysis of the typical characteristics of Science and Society initiatives. The second part reviews evaluation approaches to Science and Society programmes and projects. As with the preceding section, this section concludes with a 'mapping tool' to help provide inputs to subsequent evaluation design.

4.3. Types of Science and Society initiative

A review of the literature suggests that Science and Society initiatives can be broadly grouped into nine basic categories, as shown in Table 4-1.[9]

[9] see, for example, Edwards (2004); Gascoine and Metcalf (2001); Thomas (2002); Keogh (1999); Jarvis and Pell (2002)

Table 4-1 Types of Science and Society initiatives

Category	Distinguishing features
Large scale awareness-raising campaigns	Typically national campaigns on specific issues or generic issues but compressed into a short timescale, e.g. 'Science Week'
Public participation	Typically 'exploratory' actions intended to involve public in issues where there is significant concern, where the science is embryonic, or where there is little evidence base
Interactive events (outreach; theatre; demonstrations)	Covers a wide range of generally highly structured, small scale and typically 'one off' actions, often involving the use of existing dedicated facilities, e.g. Natural History Museum
Education and training	Covers a spectrum of possibly contrasting actions of varying scales, from large scale national programmes – for example, Science Learning Centres – through to specific targeted learning events
Ongoing profile-raising	Covers more 'ephemeral' activities such as: lobbying; popular science journalism and broadcasting; networking
Targeted access and inclusion actions	Focused on specific accessibility issues such as gender balance and low representation of black and ethnic minority communities in science professions
Policy actions	Initiatives and interventions aimed directly at shaping policy agendas and instruments, for example, consultation processes
Horizontal and supporting actions; capacity-building	'Indirect' actions intended to support and facilitate more direct initiatives. For example, subsidies and incentives to enable schools to take advantage of training programmes
Operational Reviews	Reviews, research and analysis commissioned to assess state of the art and synthesise evaluation reports and studies

As Table 4-1 shows, the factors that distinguish between the different categories identified are not always clear. However, a plausible classification approach might highlight the following discriminating factors:

- Scale – the extent to which the initiative involves the national population or is more narrowly targeted.
- Stakeholder configuration – the representation of different groups of actors and their relationships.
- Engagement – the extent to which lay groups as well as scientists and other professionals are involved in knowledge creation.
- Scenario – the actual delivery mechanisms involved, for example: direct human interaction; mass media; new technologies.
- Overall purpose – the main objective of the initiative, for example, awareness-raising or skills development, and particularly the extent to which the purpose is 'open-ended' or 'specific'.

- Specific outcomes – the concrete outcomes intended, for example, an increase in the number of girls taking science 'A' levels.

Using this initial starting point as a basis, we began to expand this typology by examining the features of the DIUS Science and Society programme itself, in comparison with another major UK programme concerned with supporting and developing Science and Society: the NESTA 'Learning Programme'. Table 4-2 and Table 4-3 show the results of this classification. In both programme cases, we identified additional distinguishing variables to those listed above. These included:

- Innovation level – the extent to which the 'science' embedded in the initiative is intended to promote a limited contribution to developing new paradigms; a modest enhancement to existing paradigm; a new dimension to an existing paradigm; a significant enhancement; a radical new paradigm.
- Pedagogic mode – the 'theory of change' and attitude and behaviour modification models incorporated.
- Level of analysis – the scale and level of the initiative itself – whether a policy, 'programme' or project (not specified in Tables 4 and 5 since both examples are programmes).
- Timescale – the duration of the action and the timeframe for expected results, outcomes and impacts.

Table 4-2 Examples of NESTA Science and Society initiatives

Name	Description	Type/ Purpose	Target group/ Audience	Specific outcomes	Innovation level	Timescale	Scenario/ Setting	Stakeholders/ Partnership
Failing to Learn	A 'Lab of Learning' for science teachers to improve their teaching skills	Education	Science teachers	Reduce professional drop-out rate	New	Medium	Residential courses	Commercial/ bureaucratic
Motivate	To get talented secondary school pupils interested in mathematics through video-conferencing	Education	Young people	Increase participation in science subjects	Modest	Short	In-school	Academic
Plant takeaway	To communicate human-plant interdependency by building a life-size automaton of a kitchen scene	Awareness	Public	Increase public interest in plants	Low	Short	Roadshow	Industrial/civic
Young Foresight	To hone the skills young people will need to work in industry, to develop their ideas into products and to bring those products successfully to the marketplace	Horizontal	Young people	Increase technical and entrepreneurial skills	Low	Medium	Outreach	Industrial
ACRISAT	To advance the educational achievements and career aspirations of black youth within the fields of science, mathematics and technology	Access	BEMGs	Improve learning outcomes	Low	Long	Community-based	Civic
Science Worlds UK	To help to establish UK's science centres as significant visitor	Horizontal	Public	Increase attendance at science centres	Low	Long	Purpose-built centre	Civic

Name	Description	Type/ Purpose	Target group/ Audience	Specific outcomes	Innovation level	Timescale	Scenario/ Setting	Stakeholders/ Partnership
	attractions							
Science Line	To provide a free science information service and to encourage young people to take up science and develop a career in it	Participation	Young people	Promote use of information service	Low	Medium	Outreach	Bureaucratic
Planet Jemma	To encourage more teenage girls to consider science-related courses and careers by creating an interactive drama about the life, loves and university career of a young first-year physics student, Jemma	Access	Young females	Build webcam audience	Significant	Long	Mass media	Industrial/ bureaucratic
Lecture List	On-line resource promoting any lecture on any science subject taking place in the UK	Horizontal	Public	Build on-line consumer base	Low	Medium	On-line	Mixed
Scottish Executive Small Grants	Grant scheme to boost science skills and knowledge in Scottish schools	Horizontal	Teachers	Get schools to apply for grants to support science teaching initiatives	Low	Short	On-line Workshops	Bureaucratic

Table 4-3 Characteristics of DIUS initiatives

Project/ Programme	Description	Audience	Overall purpose	Specific intended outcomes	Innovation level	Pedagogic mode	Stakeholders/ partnership	Timescale (outcomes)
ScienceWise	Supply grants to/commissioning of projects that help to achieve greater public confidence in and improved engagement with science and technology	Policy-makers, scientists, academics, industry, NGOs, citizens, young people	Open-ended: Help the Government and society make better choices about critical areas of new science and technology that affect people's lives	encourage collaboration; widen participation in engagement and dialogue activities; develop/disseminat e good practice	Low	Collaborative	Policy-makers, scientists, academics, industry, NGOs, citizens, young people	Long-term
Trustguide – Guidelines for enhancing cyber trust	Produce clear guidelines for the research, development and delivery of trustworthy ICT	General public; profiled subgroup (200-350 in total)	1 specific (create guidelines), 2 open-ended (disseminate/establis h dialogue). Enhance trust in ICT	Enhancing trust; encourage collaboration; develop/disseminat e GP	Low	Collaborative learning [workshops to create the guidelines]	Industry, Policy-makers, relevant researchers, general public	Short-term (guidelines) / long-term (establish dialogue)
Nanodialogues – the Nano Dialogues: Four experiments in upstream public engagement	Scoping project, to investigate the most appropriate methods of 'upstream' engagement, designed to inform decision-making. Undertaking 4 experiments to identify appropriate method for upstream dialogue	General public; decision-makers	Specific. Experiment with new methods of 'upstream' public dialogue	Identify appropriate method/Public engagement	Significant	Collaborative learning [workshops]	Scientists, policy-makers, businesses; general public	Short-term (identify methods)
Democs - deliberative meeting of citizens	Pilot project to develop and trial a card game to promote dialogue and learning for young people around	Young people	Specific – usefulness of tool to promote dialogue and learning for young people around controversial	Pilot: eval of usefulness of tool [Ultimately: Awareness-raising; knowledge	Moderate	Not clear	Policy-makers, scientists, young people	Short-term (develop game) / Long-term (raising awareness/knowled ge enhancement)

Project/ Programme	Description	Audience	Overall purpose	Specific intended outcomes	Innovation level	Pedagogic mode	Stakeholders/ partnership	Timescale (outcomes)
	controversial S&T topics (e.g. climate change, gm food)		S&T topics	enhancement				
RiskyBusiness – Becomingwise	Increase young people's awareness and appreciation of the role of risk in engagement with science through drama	Young people (5,000 students / 100 teachers)	Open-ended: Encourage participants to consider how differing perspectives on risk arise	Awareness-raising / risk management	New dimension	Transmissive / Deficit model	Policy-makers, teachers, young people	Long-term
SCWL – Science Communication Working Lunches	Provide opportunities (lunches) for communicators to enhance their skills in and knowledge of current developments in public engagement	**Profiled subgroups:** Science community / those interested in science communication [attendees of lunches <700; reports – public; summaries targeted]	Specific (lunches and publication of reports / mailing of summaries): Enhance the science community's public engagement skills	Skills / professional development; Good practice development / dissemination; knowledge enhancement	Low	Collaborative	Policy-makers, sci community, NGOs, PR, industry, interested community groups.	Short-term (24 mths)
Cxchange – Citizen x-change: sharing knowledge and linking to policy	Provide a policy-focused approach to engagement that will give non-scientists and scientists a voice on ST and other issues.	Scientists; citizens [small group, attendees of workshops]; citizen-science engagement community [large: users of new dialogue model]	**Specific** (workshops and reports of; new model for dialogue); **open-ended** (strengthen links between sci and non sci; recommendations shape BA Festivals). Provide a new model for dialogue connecting top-down and bottom-up	Knowledge enhancement / sharing; encourage collaboration; good practice development	Low	Collaborative	Scientists, policy-makers, general public,	Short-term (24 mths)

Project/ Programme	Description	Audience	Overall purpose	Specific intended outcomes	Innovation level	Pedagogic mode	Stakeholders/ partnership	Timescale (outcomes)
			approaches					
Neg – Nanotech Engagement group	Map current public engagement re nanotechnology	**Specific sub-groups**: Nanotechnology scientists; academics; NGOs; Policy-makers; the public	**Open-ended** (change thinking and acting on public engagement) bring about better and more joined-up nanotechnology engagement products	encourage collaboration; knowledge enhancement; good practice development	Significant	Researching how to engage. Will also communicate the learning	nanotechnology scientists; academics; NGOs; policy-makers	Of eval: 24 months; of effects: long-term
National Science Week	Variety of science, engineering and technology events giving people across the UK the chance to participate in science activities, experiments and discussions	SET employers, scientists, academics, science-citizen communicators, public	Open-ended. Celebrate science and its importance to our lives	Awareness-raising; knowledge enhancement; encouraging SET career; public engagement; encourage collaboration	Low	Mixed	SET employers, scientists, academics, science-citizen communicators, public, policy-makers	
The court of the rainbow king (NSW) [Life]	Aimed at KS1. Children explore mixing colours with hands-on experiments	KS1 children	Open-ended. Encourage young people to develop a curiosity about science and scientific endeavour	Awareness-raising	Low	Situated	Programme beneficiaries	Long-term

Project/ Programme	Description	Audience	Overall purpose	Specific intended outcomes	Innovation level	Pedagogic mode	Stakeholders/ partnership	Timescale (outcomes)
Galileo, genetics and the greens – perspectives on scientific thought (NSW) [the great debate]	Ten-week course exploring the development of ideas about our relationship with the natural world	General public	Open-ended. Committed to public participation and to facilitating topical discussions	Awareness-raising; knowledge enhancement; encouraging critical thinking	Low	Transmissive		Long-term
Explorers: ourselves and other animals(NSW) [Natural history museum]	Encourage young children to engage with specimens through a process of role-play	Reception and KS1	Specific. Encourage young children to engage with specimens through a process of role-play	Awareness-raising; youth engagement; knowledge enhancement	Low	Situated		Long-term
Future human competition (NSW) [Computer Science for Fun]	Schools competition involves pupils predicting how the human of 50 years from now will look, communicate, work, play and live, basing predictions on actual computer science projects	School pupils nationwide	Specific: enter competition – raise awareness / knowledge; Open-ended: engagement. Promoting computer science and the view that it is also about people	Awareness-raising; youth engagement; knowledge enhancement	Low	Transmissive	computer science-citizen communicators, teachers, school pupils	Short-term (enter competition – enhance knowledge)
UK RC for Women in SET	Review and develop a recognition scheme for 'good' SET employers; share good employment practice for women in SET	Profiled subgroups: SET employers; women in SET;	open-ended (work ongoing) / specific (carrying out particular projects). Increase the participation and position of women in science, engineering and technology	hold and disseminate information about women in SET; develop good employment practice; monitor good employment practice; co-ordinate the work of partner orgs	Low		Policy-makers, SET employers, women considering / pursuing SET careers	

Project/ Programme	Description	Audience	Overall purpose	Specific intended outcomes	Innovation level	Pedagogic mode	Stakeholders/ partnership	Timescale (outcomes)
Internet Computing	Workshop will look at opportunities and training needed to work in environments that provide information and deliver news services on-line	Women pursuing an ITEC career	Open-ended (aid employment potential); specific (learn particular info)	Skills / professional development; social inclusion	Low	Transmissive	key ITEC employers, women interested in pursuing ITEC careers, and educational institutions with experience of successful delivery of ITEC courses and skills	
Get SET Women database	Provides the media and other organisations with access to thousands of women, at various stages in their science, engineering and technology careers	Women in SET, media, science-citizen communicators,	Open ended. Improve the accessibility and visibility of female SET role models	Disseminate information; social inclusion; encourage SET career	Low	Database	Policy-makers, SET employers, females considering / pursuing SET careers	Long-term

Analysis based on publicly available information

Key lessons for evaluation design derived from this analysis of the characteristics of the NESTA and DIUS programmes are:

- The complexity and range of the objectives and purposes reflected in the programmes. The wide spectrum of objectives, ranging from hosting lunches to promote discussion about science through to national collaborative learning programmes aimed at developing new paradigms for knowledge creation, means that evaluation approaches and methods need to be flexible enough to accommodate such diversity.
- Open-endedness. Although many of the initiatives in the programmes are highly specified, with clear aims and intended outcomes, a number of initiatives, notably those at the 'public participation' end of the spectrum, reflect agendas that are embryonic and rapidly evolving. For example, 'Nanodialogues' embodies, at least in some respects, elements of a 'societal learning' agenda which, in contrast to the dominant 'deficit model' agenda, envisages a much more collaborative role for non-scientists in the shaping of knowledge, particularly knowledge that has 'civilising choices'. This poses considerable challenges for evaluation.
- Complex stakeholder groups and relationships. The initiatives typically engage a range of different groups, each of which are likely to have varying: perspectives on science; expectations of the objectives and outcomes from the initiative; variable access to power and decision-making. This suggests a need for evaluation methods and tools that can: map stakeholder needs and expectations; identify conflicts; and promote 'alignment' of conflicting positions.
- Multiple delivery modes. The 'scenarios' through which initiatives are implemented constitute a wide-ranging spectrum with widely varying attributes, encompassing: popular broadcasting media through digital information systems; outreach work; and established institutional spaces such as the Natural History Museum. This variability again implies the need for flexibility in terms of selection of data gathering methods.

4.4. Evaluation of Science and Society initiatives

An important task for the project was to consider in a systematic way how real-world Science and Society initiatives are evaluated. A key aim of this part of the research was to try to build a typology that could link particular types of initiative to particular evaluation models, methods and tools, in order to provide empirical evidence to support evaluation design recommendations. As with the review of types of initiative reported above, the approach used in this exercise incorporated:

- Literature review, including a structured search of on-line bibliographic databases, and an internet search for grey literature.
- A review of relevant specialist Journals ('Evaluation'; 'Public Understanding of Science'; 'Science Communication').
- Expert interviews.

All the data sources deployed reinforced the following conclusions:

- Although a considerable volume and range of Science and Society initiatives are developed and implemented each year in the UK and further afield, only a small proportion of these initiatives include evaluation activities.
- The quality of evaluations carried out and the methods used vary considerably.
- There is insufficient evidence to identify any clear correlation between type of initiative carried out and evaluation approach and methodology used.
- Evaluation theory and practice in this particular field is limited. Existing practices and guidelines tend to focus on promoting good evaluation design and practice in general, and on identifying appropriate questions and criteria for specific Science and Society initiatives, rather than on developing 'contingency models' to match initiative type to evaluation method.
- Despite the range of evaluation methods and techniques currently available, most evaluations used a combination of pre-test/post-test questionnaire surveys, interviews and focus groups.

Our own exercise confirmed the picture presented by previous reviews of the evaluation field. For example, in a recent review of European state of the art, Edwards (2004) concludes:

> *"although there are very many PASIs[10] taking place across Europe, only a tiny proportion are involved in any form of systematic evaluation and reporting."[11]*

This conclusion mirrors the viewpoint of many stakeholders involved in the Science and Society field. For example, the British Association, in its advice to the Office of Science and Technology, following the then OST's public consultation initiative in 2002, observed that:

> *"there is little published evaluation of activities and no systematic programme to assess, for example, which modes of engagement best support effective dialogue between scientists, the public and decision-makers". [12]*

In the light of this situation, one of the BA's recommendations to OST was to promote and carry out a systematic evaluation of "a range of activities […] exploring which activities are most engaging for particular groups of people".

It has been argued that this apparent low level of interest in evaluation reflects a lack of importance attached to public awareness of science in general by policy-makers (Edwards, 2004). However, it could also in part be explained by both a generally low level of evaluation expertise and a corresponding lack of depth in the specialist expertise required to address the range and complexity of the kinds of programmes and projects that are being developed and delivered in this field. Moreover, the range and complexity of initiatives in turn reflect the rapidly evolving nature of the domain

---

[10] Public Awareness of Science Initiatives

[11] Edwards, C (2004) Evaluating European Public Awareness of Science Initiatives, Science Communication, Vol. 25 No. 3, March 2004 260-271, Sage

[12] 'Science and Society': advice to the Office of Science and Technology, British Association, 2002, London

and the fact that scientific knowledge is frequently unstable and contested. This, it could be argued, makes it difficult, if not prohibitive, for a ‘contingency-based’ evaluation approach to develop, based on matching evaluation models and methods to type of programme and project. As a result, the limited body of evaluation theory, practice and guidelines in this field tends to focus on promoting good evaluation design and practice in general terms and on identifying suitable evaluation questions and criteria to match the objectives of particular Science and Society initiatives. For example, the evaluation model and guidelines developed by the Committee for Public Awareness of Science (COPUS) in 1996 sets out six types of questions evaluations need to address and three types of method that are required to gather data to answer these questions. The types of questions cover:

- Audience and target groups. Which are the intended audiences and does the intervention adequately cover them?
- Reach and penetration. How large an audience is expected and to what extent were these expectations realised?
- Experiences. What experiences are target groups expected to gain and were these realised?
- Education. What are target groups expected to learn about science and were these learning goals achieved?
- Attitudes. Was the audience’s attitudes to science expected to be changed and in what ways did this change occur?
- Follow-up. What post-event behaviours were anticipated (e.g. join a scientific society, do projects in the classroom or at home)?

The recommended methods to gather data on these questions cover surveys, questionnaires and focus groups.[13] Successive iterations of these types of guidelines, including the evaluation guidelines developed jointly by OST and the UK Research Councils in response to the British Association’s 2002 Report, have similarly focused more on the importance of implementing common evaluation activities and asking the right evaluation questions rather than trying to match the characteristics of different types of programme or project to particular evaluation methods.

Despite this strong evidence that a contingency-based approach to evaluating Science and Society initiatives has not been developed, and is likely to be of limited value in any case, our review, taking into account similar studies and exercises,[14] nevertheless allows a picture to be built up of the broad range of the typical evaluation questions and criteria used and the ‘core‘ evaluation approaches and methods that tend to be used across the range of Science and Society initiatives identified in section 4.3 above. This is summarised in Table 4-4.

[13] So did it work? Evaluating public understanding of science events. COPUS, 1996, London
[14] See, for example, Edwards (2004)

Table 4-4 Science and Society initiatives and evaluation methods.

Type of initiative	Typical evaluation questions and criteria	Typical evaluation methods
Large scale awareness-raising campaigns	What kinds of people changed their attitudes towards science and in what ways?	Cross-sectional surveys Longitudinal (cohort) studies Science literacy scales
Public participation	How can public anxieties about nuclear power be productively harnessed to develop sustainable energy?	Citizens' juries Developmental evaluation Focus groups
Interactive events (outreach; theatre; demonstrations)	How many and what type of people attended the event? How engaged was the audience? In what ways did participants' views of science change?	Exit polls Quota sample Analysis of attendance records Observation interviews
Education and training	The number of high school students completing science courses Movement in the salary levels of scientists and technologists	Statistical analysis Questionnaire surveys Interviews
Ongoing profile-raising	To what degree and in what ways is science covered in the popular media? What contribution does profile-raising investment have to science policy and improving the knowledge base	Content analysis of sample of newspapers Citation analysis of academic journals
Targeted access and inclusion actions	Have the proportions of black and ethnic minority students achieving science degrees increased?	Statistical analysis Questionnaire surveys
Policy actions	Has the implementation of the consultation exercise created new partnerships?	Focus groups Documentation analysis
Horizontal and supporting actions	How many schools have taken advantage of subsidies for Science Learning Centres?	Statistical surveys Documentation analysis
Operational Reviews	Which public engagement approach is most cost-effective?	Process evaluation Cost-effectiveness analysis

Table 4-4 shows:

- Most evaluations deploy a limited set of common evaluation methods based on surveys and interviews.
- Some types of initiative lend themselves to particular types of evaluation methods. For example, operational reviews, aimed at assessing the effectiveness of how an initiative is delivered and managed, would typically focus on a process evaluation approach, combining analysis of documentation and interviews with key stakeholders.
- However, the key to relevant and effective evaluation is more directly related to formulating the objectives and purposes of the evaluation, in line with those of the initiative itself, and then identifying the key evaluation questions that need to be addressed and the criteria through which the evaluation can assess whether the purposes and objectives have been fulfilled. Selection of appropriate evaluation methods and techniques is shaped to a large extent by these questions and criteria, rather than the attributes of the initiative directly.

With this in mind, we present in Table 4-5 below the mapping tool for the proximal level which is designed to help evaluators understand and articulate how an initiative is constructed on the ground and how this relates to evaluation design.

4.5. Mapping tool: proximal level

As with the broader societal level, understanding how Science and Society initiatives are put together 'on the ground' will help the evaluation to specify, for example:

- The 'object' of the evaluation (what is the 'unit of analysis' to be evaluated?).
- The purposes and scope of the evaluation questions that need to be asked and the criteria that need to be used to draw conclusions.
- The perspectives and expectations of stakeholders that need to be taken into account.
- The 'spaces' in which data can be collected and the opportunities and limitations these impose for the evaluation.

The results of the mapping exercise are intended to provide inputs to stages 4 and 5 of the evaluation process, as set out in Sections 5 and 6.

Table 4-5 **Mapping tool – proximal level**

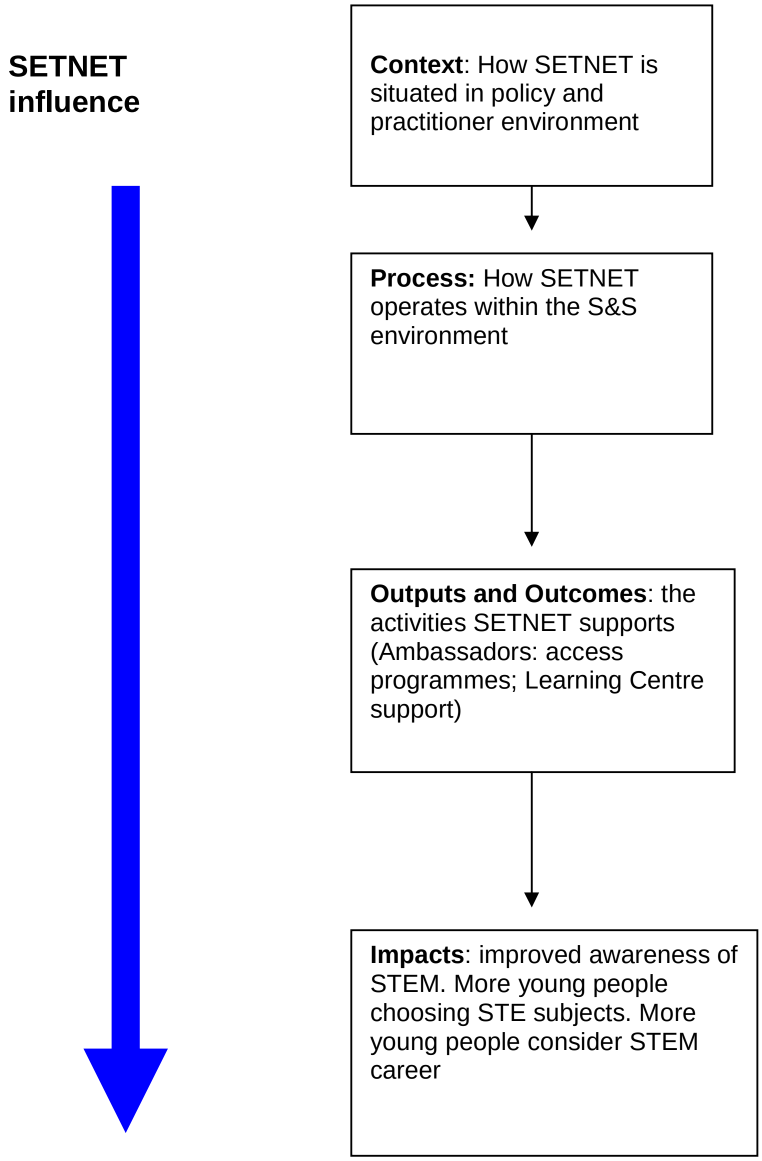
Dimension	Data required	Inputs to evaluation
Specific purposes	Outputs, outcomes and targets expected. Open-endedness v specification	Identifying the purposes of the evaluation, e.g. descriptive; analytical; operational Specification of evaluation questions and criteria, e.g.: How many people attended the 'science event'?
Scale	Coverage of the initiative: national; regional; local	Overall methodology Scale of evaluation (e.g. sample size; locations for evaluation work)
Stakeholder configuration and engagement	Stakeholder analysis: perspectives and expectations	Identifying the purposes of the evaluation, e.g. developmental; consensus-building Clarifying kind of evaluation evidence that is acceptable
Scenario	Delivery mechanism for initiative	Clarify data collection opportunities and constraints
Innovation level	Degree to which 'science' is stable or contested Innovation of 'social' dimension	Identifying the purposes of the evaluation, e.g. developmental; consensus-building Clarifying questions and criteria
Analysis level	Type of initiative: policy; single project; programme 'Architecture' of programme	Resources required for evaluation Selection of methodology and methods (e.g. case studies for programmes)
Timeframe	Duration of the action and the timeframe for expected results, outcomes and impacts	Specifying evaluation life cycle; key activities and milestones
Pedagogic mode	Mode of pedagogy utilised, e.g. 'Deficit' model Lay representation Collaborative learning	Specification of evaluation questions and criteria

4.6. Delivery chains and distance from beneficiaries

Most commissioners of evaluations are interested in the impacts of Science and Society initiatives, as achieving impact is the raison d'être of initiatives. In most initiatives, there is an expectation that they will promote some change in human behaviour, whether through: changing the public's attitudes to government policy; improving science teachers' classroom performance; or encouraging girls and young women to take up a career in scientific research. This expectation often reflects two underlying misconceptions: that different types of initiatives will have similar effects on human behaviour and that the influence of the initiative on these effects remains relatively constant throughout the duration of the initiative. However, the study has shown firstly that Science and Society initiatives broadly represent a particular 'delivery chain', reflecting the degree of their direct engagement with 'end users' such as citizens, teachers and students. Correspondingly, their capacity to promote changes in the behaviours of these end users will vary according to their proximity to users. Secondly, the study showed that this capacity is itself variable throughout the life cycle of an initiative. It changes over time for a number of reasons, for example, through changes in how the initiative is managed or modifications to the initiative's objectives. These factors significantly affect the strategies that need to be adopted to evaluate the initiative.

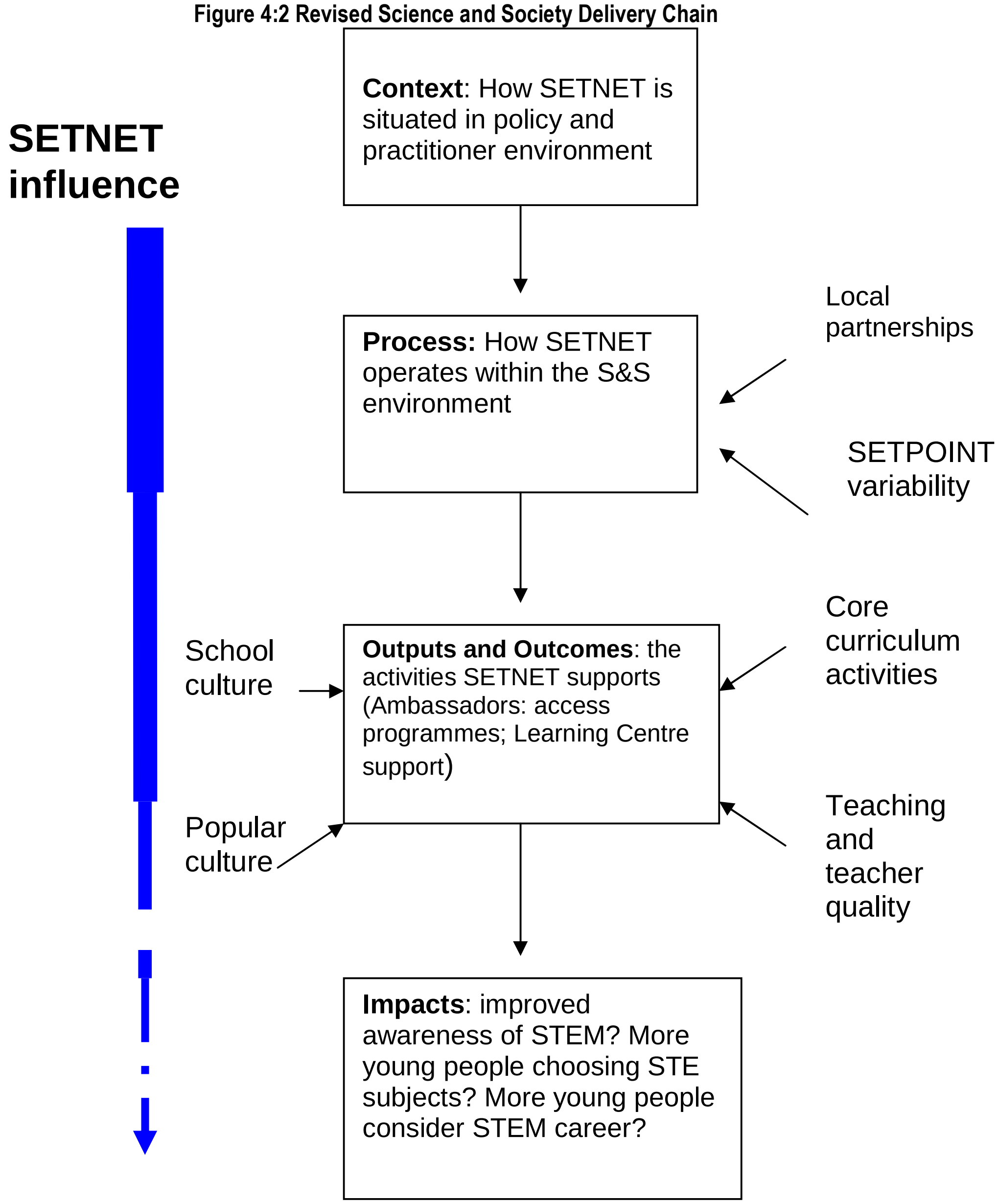
The case of SETNET, which was selected to pilot test our Evaluation Framework, is a good illustration. An important assumption underlying our initial evaluation approach was that SETNET activities have direct effects at the end of the delivery chain, i.e. in terms of affecting young people's attitudes to STEM subjects, and their decisions involving subject choices and career planning. Another assumption was that the influence SETNET brings to bear on these 'end user' outcomes is relatively uniform. An important part of the evaluation, therefore, looked at the outputs, outcomes and impacts of SETNET's operations. It reviewed the activities carried out by SETNET, such as the Science and Engineering Ambassadors (SEAs) programme; the projects aimed at improving access to STEM subjects for 'hard to reach' groups; and SETNET project management of the establishment and running of the new STEM Support Centres. This part of the evaluation also examined the type of outcomes associated with these activities, for example, by observing how young people engaged with a SETNET-supported 'science event' and the benefits they attributed to their engagement. Finally, the evaluation aimed to assess the impacts associated with these outputs and outcomes in two main ways: through a survey of teachers (exploring how curriculum enrichment programmes including those supported by SETNET affects teachers' skills development, quality of teaching and learning impacts) and a students' survey (supported by focus group discussions) examining how working with SETNET affects young people's attitudes to science.

Figure 4:1 Assumed Science and Society Delivery Chain



As the evaluation progressed, however, it became obvious that our initial definition of the 'object of evaluation' needed to be re-assessed. It became clear that SETNET's main function was to act as a mechanism to support the formation and development of partnerships, which in turn enabled STEM support services to be provided to schools. This meant that SETNET was less directly influential in terms of its effects on the behaviours of teachers and students than those actually providing such services in schools. The outputs, outcomes and impacts associated with SETNET's role are more concerned directly with, for example, the nature, quality, effectiveness and sustainability of these partnerships rather than with the perceptions and decision-making strategies of young people. Thus SETNET primarily intervenes at a point much further up the 'delivery chain' rather than at the end of that chain, i.e. at the interface with partnerships rather than in the classroom. As Figure 4:2 suggests, its influence in terms of direct impacts on young people's perceptions and behaviour is weaker at the level of the classroom and stronger at the level of the partnership – where it is engaging with entities that more directly provide STEM activities having a bearing on how young people think. This situation is further complicated by the variability of the curriculum enrichment and other STEM support actions developed and implemented through SETNET. A key finding of the evaluation was that the 53 sub-regional SETPOINTs supported by SETNET, through which direct engagement with schools takes place, vary considerably in terms of effectiveness, efficiency and quality of outputs. This means that SETNET's influence (and ultimately its impact) is not constant but will vary in line with SETPOINT variation.

Figure 4:2 Revised Science and Society Delivery Chain



Moreover, by the time SETNET's contribution and influence works its way through to the end of the delivery chain, its impact is mediated and filtered through an ever widening array of additional factors that also make a contribution to shaping the ways young people think and act about science and related subjects. These include:

- the organisational structure of the schools
- the culture of the school
- the quality of teaching and the effectiveness of teachers
- the influence of external factors affecting young people's perceptions and attitudes, particularly popular culture

The main lesson for evaluation design is that choices on evaluation approach and methods need to be informed by the type of initiative being evaluated and its position in the end user 'delivery chain'.

Part II: Building an Evaluation Framework for Science and Society initiatives

5. THE EVALUATION FRAMEWORK: BUILDING BLOCKS

5.1. What this section is about

This section provides a template for evaluating Science and Society programmes and initiatives. It covers: design principles; the evaluation life cycle; how to apply the design principles to the characteristics of the initiative being evaluated; developing evaluation questions and criteria; choosing the right data collection methods and techniques; implementing the evaluation; and utilising and disseminating the evaluation results. In summary, this section guides the evaluator through the different stages of the evaluation process.

This section integrates the results of the literature reviews, analysis and classification of Science and Society initiatives and review of evaluation approaches within an evaluation design framework to produce a template for evaluating Science and Society initiatives. The following considerations, drawn from these results, have particularly shaped the development of this template.

5.2. Science and Society domain characteristics

Here, we summarise the key characteristics of the Science and Society domain which will inform evaluation design. For a fuller discussion see section 3 above.

The nature of scientific knowledge. As we have argued, science is continually being constructed and re-constructed. This happens in episodic ways – sometimes precipitated by a particular crisis – and much of the science that enters into, and is engaged with, the public domain is stable and relatively uncontested. However, the value, relationship to social life and the expectations associated with this knowledge can be highly contested. Moreover, from time to time, 'new science' emerges which is itself embryonic, rapidly evolving and inherently contested. This engenders a 'context of meanings' between different constructions of science which needs to be understood and handled in evaluation.

Environmental and organisational turbulence. Science and Society essentially operates in 'change' mode. The objectives of initiatives are predicated on promoting change in public awareness; in attitudes; in behaviours; in skills; in the management of policy and programmes. Change operates across a range of structures, processes and levels within and outside the environment of the initiatives, at both the proximal and distal levels. Organisationally, the initiatives work in what systems theorists like to call a 'turbulent environment' (Trist and Emery, 1966; Miller, 1991). They engage with and respond to demands from a range of stakeholders, and do so against a constant backdrop of change. In one important sense, Science and Society programmes are themselves instruments of change, seeking to promote the engagement of scientific theory, research and practice more broadly within social life. At another level, their mission is often to support policies that consider learning as itself a 'transformative' process, for example, in the linkages between lifelong learning and social inclusion.

Societal learning. In turn, there is currently strong evidence that citizens are demanding not only more information relating to how science shapes their lives but more of a say in the policies and practices through which science becomes embedded in society and through which it provokes real changes in people's everyday lives. It is, therefore, likely that an emergent new mode of social engagement with science, akin to Giddens' idea of 'dialogic reflexivity' (see Section 3.3 above), will precipitate new forms of public participation based less on the passive 'information seeking' epitomised by the prevailing 'deficit' model of public engagement but more on 'societal learning'. This in turn requires a new evaluation modality centred on a developmental approach in which evaluators actively engage as collaborators with stakeholders in the co-production of knowledge.

Pluralism and power relations. Public awareness of science programmes and other Science and Society initiatives typically involves complex stakeholder configurations, members of which are likely to hold different perspectives on science, on the objectives and outcomes expected for the initiative, and variable access to power and decision-making. This suggests a need for evaluation methods and tools that can: map stakeholder needs and expectations; identify conflicts; and promote 'alignment' of conflicting positions.

Multiple visions, purposes and scale. Different social constructions of science reflect a multiplicity of perspectives and positions on how science and the social world should engage with each other. Moreover, the 'scenarios' through which initiatives are implemented constitute a wide-ranging spectrum with widely varying attributes. This variability again implies the need for flexibility in terms of selection of data gathering methods

5.3. Design principles

In response to these considerations, the proposed Evaluation Framework embodies the following design principles:

Mutli-dimensional and multi-methodological. The variability and complexity of the domain precludes the development and implementation of a 'one size fits all' framework. Furthermore, the research strongly suggests that a 'contingency-based' approach, involving matching the attributes of a particular initiative to a given configuration of evaluation methods and tools, is neither feasible nor desirable. The evaluation of complex intervention processes, therefore, implies an 'open systems' perspective which gives recognition to the organisational and political processes within which it is embedded (Chen, 1990). In turn, the need to reflect different constructions of what science is, and what its social role should be, implies some reference to a 'constructivist' approach (Guba and Lincoln, 1992). And this is in addition to approaches geared towards measuring success and effectiveness, such as 'experimental' approaches which typically compare a control group and an intervention group. Ultimately, this requires the use of methodological triangulation. The evaluation needs to adopt different methods and collect a range of data, reflecting the different perspectives of the different stakeholders. In practice, this means: firstly, a stakeholder analysis of the key constituencies involved in the initiative; secondly, multiple sources of data; thirdly, a multi-evaluation methodology.

Multiple purposes. As discussed above, Science and Society initiatives encompass a wide spectrum of purposes. Therefore, the Evaluation Framework supports a similar wide rage of evaluation purposes, including:

- Ascertaining 'what works' and why (focusing on outputs, outcomes and impacts).
- Operational improvement (focusing on contributing to effective management of the initiative).
- Learning lessons about similar initiatives (focusing on transferability of results to other programmes and initiatives).

Change-focused and change-oriented. A focus on identifying the 'theory of change' underpinning the purposes and objectives of the initiative and the corresponding 'logical model' that is developed to implement and manage the desired changes.

Multiple stakeholders and perspectives. However, one of the drawbacks of using 'logical models' in evaluation is that they tend to overestimate the degree of consensus amongst stakeholders around perceptions of 'mission', objectives, desired outcomes and impacts. This means that the 'object', or focus, of the evaluation is itself a moving target. In turn, the mission and remit of the initiative could cover a wide and eclectic spectrum of projects and stakeholders. The Evaluation Framework, therefore, attempts to be able to reflect the different positions, perspectives and backgrounds of different stakeholders and, more importantly, their sometimes differing constructions of science and reality.

Future orientation and foresight dimension. The evolving, contested and multi-dimensional nature of science within its social context requires the Evaluation Framework to be receptive to predicting and modelling future development trajectories.

5.4. Evaluation life cycle

Just as the initiative being evaluated has a life cycle and progresses through different stages, so does its evaluation, and the methods and tools appropriate for each stage of the evaluation differ. The Framework covers the key stages of the life cycle of an evaluation. This is the sequence of activities that should be followed to promote a successful evaluation. The key stages of the life cycle of an evaluation are:

- Stage 1: Mapping tasks, including:
    - Understanding the object of the evaluation – the 'thing' that is to be evaluated.
    - Confirming the purposes of the evaluation and what objectives it is expected to achieve.
    - Identifying the key stakeholders and what they expect the evaluation to achieve.
    - Identifying the main evaluation questions that need to be addressed and the criteria for success that will be used to draw conclusions.
- Stage 2: Method tasks, including:

- Choosing the data collection and analysis methods and tools that will most effectively realise the purposes and objectives of the evaluation.
- Pragmatic considerations of the resources (time, budget, skills) available for the evaluation.

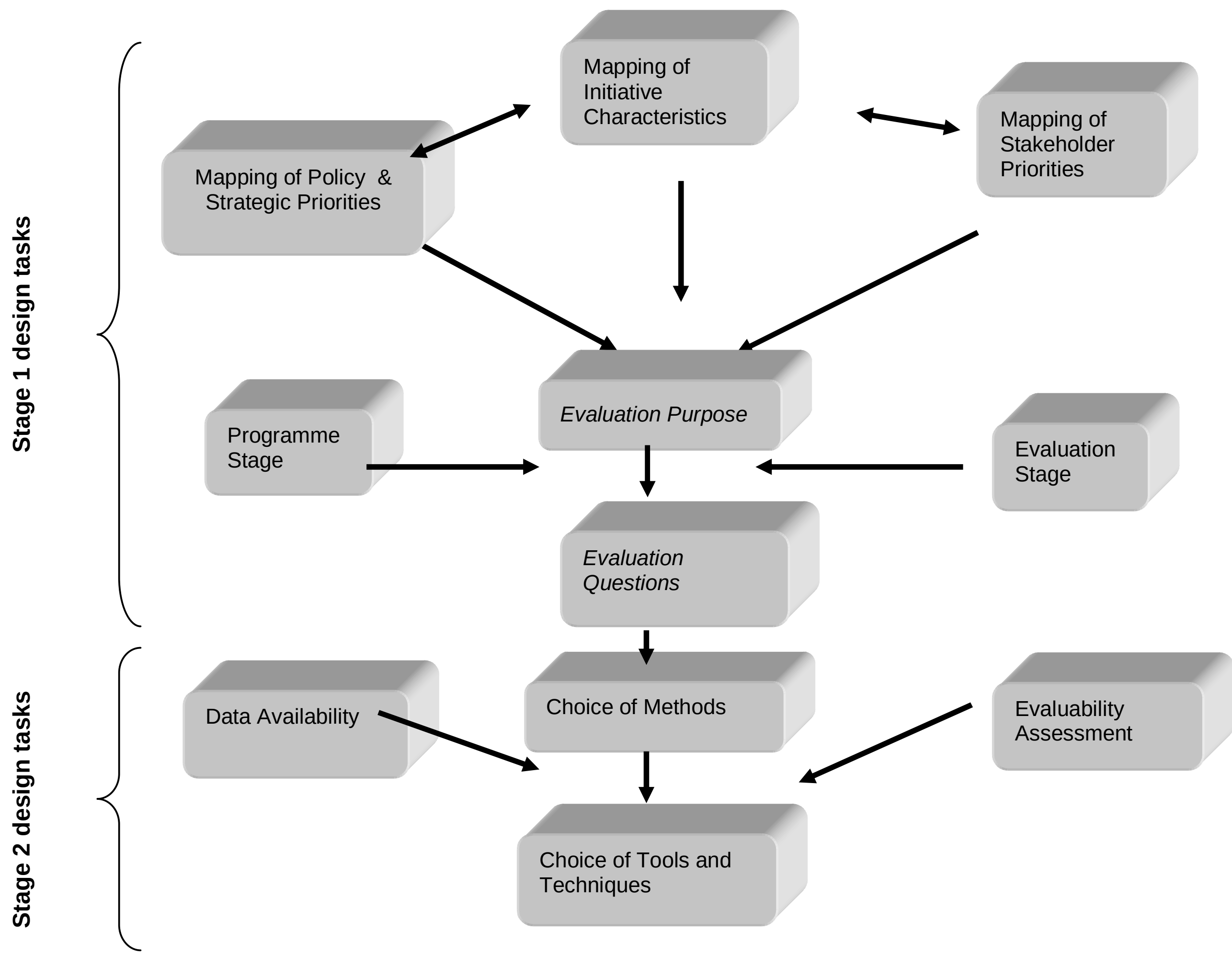
- Stage 3: Implementation and analysis.
- Stage 4: Reporting and dissemination.

There are two important caveats to bear in mind when thinking about the life cycle of the evaluation:

- The life cycle does not necessarily follow a linear sequence of events. In practice, there are likely to be several iterations of 'mini-cycles'. For example, an initial evaluation design, including the data collection tools, may be pilot tested and then refinements made to the design.
- Some evaluation methodologies, for example, those based on developmental models, may involve variations on the standard life cycle model.

Figure 5:1 is a visual representation of the key components to be considered for successful completion of stages 1 and 2 of the evaluation life cycle. In the following sections we describe these components in some detail.

Figure 5:1 Key components of evaluation design (adapted from the MEANS Handbook)



5.5. Stage 1: Scoping and mapping design tasks

5.5.1. Mapping Science and Society initiatives and their environment

The starting point for evaluations should be a 'scoping exercise'. This will enable the evaluation team to: gain an understanding of the 'distal' and 'proximal' environments in which the Science and Society initiative operates; and carry out an 'audit' of relevant data sources and contacts. The main objective of this scoping exercise is to provide the basis for developing an evaluation design. Associated objectives include:

- To deepen understanding of the initiative and the policy and practice environment within which it operates.
- To identify key stakeholders and relevant data sources.
- To clarify the purpose(s) for undertaking the evaluation.
- To promote 'sense making' amongst stakeholders; develop a shared vision and commitment to the evaluation; and lay the ground for collaborative working.

In Part I above, we described how we developed a mapping tool designed to support these objectives, the achievement of which are needed in order that the stage 1 design task of developing appropriate evaluation questions can be completed. The tool, which in Part I was divided into a distal mapping tool and a proximal mapping tool, is presented in Table 5-1. A fuller description of the various dimensions of the tool is given in Part I above, but a brief discussion is given here.

The tool uses 'cultural logic analysis' – this can be defined in terms of four key elements which reflect the 'vision' of an initiative in relation to:

- What it sets out to do in terms of the aims and objectives ascribed to it by key stakeholders (universalisation).
- How coherent the vision is, in terms of the extent to which it is shared by stakeholders, (closure).
- The practical choices made to realise the vision and its objectives (specification).
- The emphasis on and capacity for learning from the innovation or initiative, and hence its capacity to evolve and adapt in response to external and internal influences (situational change).

The cultural logic of a Science and Society initiative may reflect, for example, particular notions of national economic and social objectives and the role that science plays in these. These may emphasise: human capital constructions, about driving the economy forward through science; social capital constructions, focusing on the intrinsic value of science and its contribution to promoting individual and social development; and humanist (needs-focused) constructions, focusing on human and social problems, and the tools science can offer to solve them. Understanding how Science and Society interventions are constructed in terms of these elements enables an assessment of the 'goodness of fit' between the construction of science embodied by the initiative and that of the environment within which the initiative will function.

Table 5-1 Science and Society initiative mapping tool

	Dimension	Data required	Inputs to evaluation		
Distal dimensions	Cultural logic (universalisation)	Overall vision and purpose of initiative Goals and aims	Identifying the purposes of the evaluation, e.g. descriptive; analytical; operational Specification of evaluation questions and criteria, e.g. Cultural relevance – in what ways does the initiative support or run counter to how this aspect of science has been viewed over time?	Theory of change	Delivery chain
	(closure)	Stakeholder representation and power relationships Degree of consensus / contention of subject matter Level of current knowledge	Identifying the purposes of the evaluation, e.g. developmental; consensus-building Specifying the evaluation mode, e.g. external / objective; self-evaluation; peer support Specification of evaluation questions and criteria, e.g. Accessibility – does the intervention exclude particular groups?		
	(specification)	Specification – mapping the methodological and operational choices made	Specification of evaluation questions and criteria, e.g. User-friendliness and usability – are users satisfied with the vehicle chosen to deliver the intervention?		
	(situational change)	How learning and transferability of knowledge is reflected	Identifying the purposes of the evaluation, e.g. developmental; consensus-building		
	Initiating circumstances	Factors influencing origination of the intervention	Identifying the purposes of the evaluation, e.g. descriptive; analytical; operational		
	Policy agenda and focus	Link to specific science policies Relationship and integration with other policy areas	Specification of evaluation questions and criteria, e.g. in what ways has the initiative supported science policy? Specification of evaluation questions and criteria e.g. what value has this initiative contributed to employment policy?		
	epistemological focus	The kinds of 'evidence' that are considered credible	Helping to decide on the kinds of methods and techniques used to gather data		
Proximal dimensions	Specific purposes	Outputs, outcomes and targets expected. Open-endedness v specification	Identifying the purposes of the evaluation, e.g. descriptive; analytical; operational Specification of evaluation questions and criteria, e.g. How many people attended the event?		
	Scale	national; regional; local	Overall methodology; scale of evaluation		
	Stakeholder configuration and engagement	Stakeholder analysis: perspectives and expectations	Identifying the purposes of the evaluation, e.g. developmental; consensus-building Clarifying kind of evaluation evidence that is acceptable		
	Innovation level	Degree to which 'science' is stable or contested Innovation of 'social' dimension	Identifying the purposes of the evaluation, e.g. developmental; consensus-building Clarifying questions and criteria		
	Analysis level	Type of initiative: policy; single project; programme 'Architecture' of programme	Resources required for evaluation Selection of methodology and methods (e.g. case studies for programmes)		
	Timeframe	Duration of the initiative and the timeframe for expected results, outcomes and impacts	Specifying evaluation life cycle; key activities and milestones		
	Pedagogic mode	Mode of pedagogy utilised, e.g. 'Deficit' model or collaborative learning	Specification of evaluation questions and criteria		

In the case of Science and Society, the overall purpose of the initiative may be open-ended or tightly bounded, broad or specific. The aims and specific effects that the Science and Society initiatives seek to realise are diverse. Common themes include: raising awareness of science in general; promoting science as an attractive career option; involving in science activities those people and groups that have not traditionally participated in such events; contributing to the development of scientific skills; and seeking to influence individual and institutional behaviour or policy. The aims of the programme or project translate into the goals or objectives which the project, if successful, will achieve. Those with tightly defined aims are easier to operationalise and lend themselves to objective-driven evaluations, while less bounded aims usually require more deliberative or dialogical approaches, or considerable work to develop robust theories of change and appropriate proxy indicators. Increasing the number of children from black and ethnic minority communities who visit a particular museum, for instance, is a much more tightly defined aim than seeking to change young people's relationship to science, and it is, therefore, more straightforward to measure the hoped for change.

It is useful to distinguish between three different kinds of 'effects':

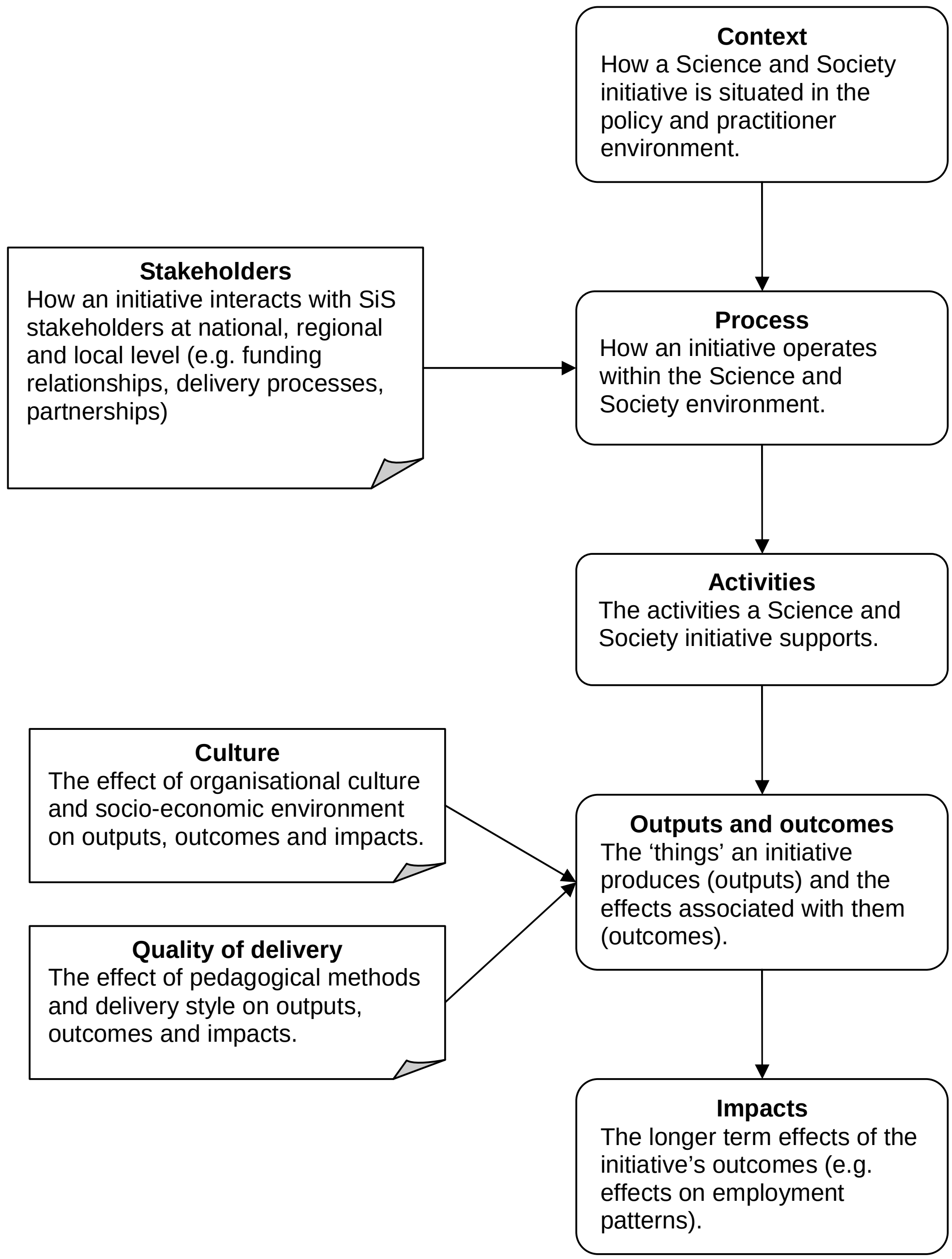
- ***Outputs*** – the things an initiative produces (for example STEM teaching resources).
- ***Outcomes*** – the immediate effects associated with the use of an initiative's outputs (for example, the effects of a 'teacher fellowship' scheme on science teachers' skills).
- ***Impacts*** – the longer term effects of the initiative's outcomes (for example, the effects of physics teachers' improved skills on their pupils' achievements and career choices).

These three types of effects are often confused and are often mixed together. To a large degree, the three types equate to the timescales of an initiative. Within the context of Science and Society, 'impacts' frequently imply changes in the behaviour of 'end users' or beneficiaries, such as citizens, teachers and students. These changes take place over a relatively long period of time, and evaluations of initiatives seeking to effect such changes should ideally be conducted over a long period of time. Similarly, 'outcomes' usually imply shorter-term changes, for example, changes in attitudes associated with a public awareness campaign. An evaluation of the outcomes of a 'one-off' event, like a road show, would normally be limited to eliciting immediate feedback from the audience about whether and in what ways their perceptions of science had changed. Longer term evaluation methods would be required to determine if the impact of the event was lasting. One would not expect a young person's relationship to science to alter overnight and yet, in a short period, you may convince a child or their parent that a trip to the science museum is worthwhile. Programmes or projects with long-term or broad aims may need to be broken down into a series of steps or mid-term objectives which can be used to see if the initiative is on track. 'Proxy measures' for the desired impact measure can be employed: while you can't directly measure a young person's relationship to science, you can measure their choice of A-level subjects, their knowledge of science and the amount of science related activity they engage in during their free time.

A further consideration is that the exact nature of the expected outcomes and impacts of an initiative also depend on the type of initiative being implemented. For example, a programme with the objective of developing the funding infrastructure necessary to support local partnerships in order to create and deliver STEM initiatives may not have a 'direct' impact on changing school students' career aspirations, but may have significant impacts on the national STEM skills base further on in time. This brings into play another feature of Science and Society initiatives that must be taken into account in designing evaluations: the position they occupy in the 'delivery chain'. Broadly speaking, it is difficult to evaluate the impacts of an initiative on 'end user' or beneficiary behaviours the further removed it is from directly engaging with these users. This is mainly because of the effects of 'intervening variables'.

Figure 5:2 suggests a skeletal outline of a delivery chain for Science and Society initiatives which can be built upon as information about the evaluation object is mapped. Evaluators should be prepared for the 'real' chain to be much more complicated than they first assumed it would be and that the distance between the initiative being evaluated and its intended impacts may well increase. Testimony to this is the case of our pilot evaluation using the Evaluation Framework, which is discussed in Part I above.

Figure 5:2 Model of generic Science and Society initiative delivery chain



Like the delivery chain, the theory of change and associated logic model combines the distal and proximal dimensions of the initiative. It represents the programme or project, linking its contexts, assumptions, inputs, intervention logics, implementation chains and outputs and outcomes. These models and theories can then be tested by the evaluation, increasing our knowledge of how and in which circumstances initiatives work. The models can be relatively simple or more complex (such as the context-mechanism-outcome configurations of Realist Evaluation[15]). It is likely that a complex logic model or theory of change will be required in order to successfully evaluate Science and Society initiatives.

A possible high-level theory of change to describe the Science and Society agenda as a whole might be:

> *There is a lack of trust in science and scientists that is based partly on a lack of public understanding and partly on a lack of communication and engagement – the latter of which fuels the lack of understanding. Given the lack of understanding in science among the public, there needs to be systematic, ongoing and institutionalised engagement between the public, policy-makers and the scientific community – including individual scientists and their institutions. This engagement or dialogue will inform the public, help scientists communicate and address public concerns, reduce uncertainty and hence increase public trust in science. In this way the legitimacy (or licence) of science will be enhanced and decision-making about science policy will become easier.*

A high-level theory such as this can be worked up at much more detailed level, capturing how the initiative is believed to function as a whole.

Science and Society initiatives have many stakeholders. These typically include: policy-makers; programme or project commissioners; programme or project managers; target audience; general public; trade and industry leaders; and professional scientific organisations. The arrangements between stakeholders who are formal partners in an initiative can vary between top-down methods of command and control with a clearly defined lead partner and more collaborative or pluralist arrangements. The various stakeholders will have differing, and potentially conflicting, priorities for the evaluation. Policy-makers tend to see evaluation as providing not only information about the success or failure of their policies but also as a tool for ensuring accountability and, assuming the programme is a success, justifying their decisions. In addition to this, programme and project managers will be interested in information which will help them to run the initiative more effectively and efficiently on the ground. Professionals and special interest groups may see the evaluation as an opportunity to enhance their professional knowledge, standing or policy position. The target audience of the initiative or general public may see the evaluation as a tool for shaping the initiative to best suit their needs. Similarly, policy and strategic priorities will also come into play.

The scale of an action, initiative or intervention can be measured in very simple terms: small, medium or large, and gradations of these three levels (very small, very large and so on). In broad terms, therefore, they are typically distinguished by:

- Single events – such as a local ‘science fair’.
- Multiple events (small scale) – such as a series of science fairs.

[15] See Pawson and Tilley (1997).

- Projects (medium scale) – of more extensive scope and objectives than ‘events’.
- Programmes (large scale) – incorporating combinations of events and / or projects, with multiple stakeholders and multiple objectives.

Initiatives may be innovative, such as Planet Jemma,[16] (outlined above in section 4) or rely on well-established methods, such as Science Week. This dimension also relates to that of the stability of the knowledge base in that radical departures from the status quo are likely to rely on a fledgling knowledge base, while traditional techniques are more likely to rest on established foundations. Innovative initiatives tend to require a more exploratory type of evaluation.

The pedagogic mode is similarly related to the distal dimensions of cultural logic and epistemological focus. Initiatives may subscribe to the ‘deficit model’ where a lack of public understanding of science is assumed as the root of any mistrust and can be remedied through straightforward communication and learning. Alternatively, they may adopt a more collaborative approach to learning, where both the public and the ‘experts’ expect to learn about Science and Society from the initiative.

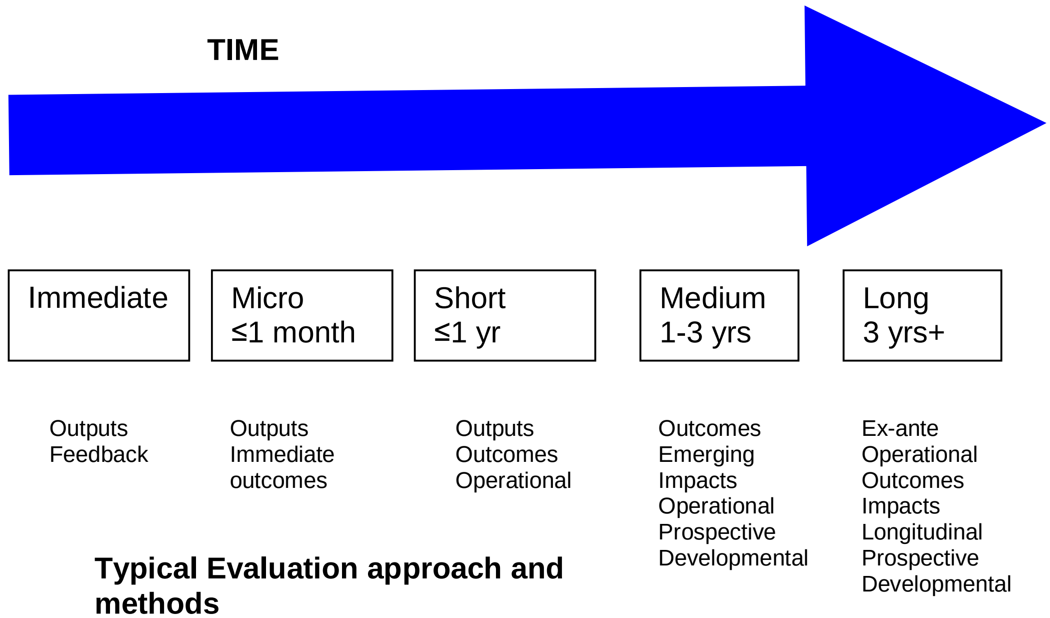
This project has suggested that there are five broad approaches or combinations of approaches that can be used to support the various purposes of evaluations of Science and Society initiatives:

- ‘Effects’ approach. This has a ‘summative’ purpose and is intended to establish what works, for whom, under what conditions.
- ‘Planning’ approach. This aims to support the scoping and design of the initiative prior to its implementation.
- ‘Developmental’ approach. This aims to support the evolution of the initiative through, for example, improving its capacity as it develops.
- ‘Operational’ approach. This aims to review and support the implementation of the initiative by focusing on its processes, for example, management structures.
- ‘Accountability’ approach. This aims to assess the degree to which and in what ways resources have been utilised by the initiative.

These different evaluation approaches are broadly associated with particular phases in the life cycle or ‘moments’ of a programme or initiative, as is shown in Table 5-3.

[16] Planet Jemma is a 14-episode BAFTA nominated on-line multimedia drama designed to excite and maintain the interest of teenage girls in studying physical sciences at university. It includes a Web video, personalised email, quizzes, message boards and Flash animation.

Figure 5:3 Relationship between initiative timescale and evaluation approach



An evaluation's purpose and requirements are likely to vary at different stages of the initiative's life cycle. Evaluation can be utilised to assess the need for an initiative or inform its design, which is usually described as ex-ante evaluation (see Annex I: Innovative Methodology). Evaluation can also take place mid-term, in order to see if initiatives are on target and to refine their implementation, or ex-post, in order to see if they 'worked'. Appropriate questions and methods will clearly vary depending on the timing of the evaluation.

5.5.2. Developing evaluation questions and criteria

Mapping the evaluation object and purposes creates the necessary conditions in which evaluation questions and criteria can be clearly defined. Any one evaluation, however, may embody a number of, sometimes conflicting, purposes. Each purpose shines the spotlight of the evaluation on a slightly different part of the initiative and will require different evaluation questions to be asked. Purposes of planning and efficiency relate to questions around value for money and whether or not the initiative constitutes the best use of public money: are there alternative programmes which would use resources more effectively?

Methods often used to answer these questions include cost-benefit analysis or objective-driven techniques. Purposes of accountability are also concerned with the efficient and effective use of resources and lead to questions such as how successful the programme has been and whether or not it has met its targets and had its desired impact. Evaluation for accountability purposes can resemble the work of auditors by using indicators to measure against relevant standards and benchmarking against similar initiatives. Evaluations designed to improve implementation will ask questions about the functioning of management and partnership arrangements, if the initiative is being delivered on schedule and as intended, and if it is targeting its desired audience correctly. Evaluations concerned with implementation are often formative evaluations which take place alongside the delivery of the initiative in order that information can be fed back in real-time to programme or project managers. They tend to be descriptive, painting a picture of processes and interim outputs or outcomes.

Evaluations seek the answers to questions about whether or not an initiative is working and the reasons for the level of success achieved. They may also try to draw conclusions that could apply to similar initiatives. Experimental methods are associated with answering ‘what works?’ questions, but are often only appropriate in very limited circumstances. Realist methods are more readily applicable to programmes or projects which employ complex mechanisms to achieve their aims. Information may also be sought from outside of the evaluation via a synthesis of existing knowledge on comparable initiatives.

When the purpose of an evaluation is to strengthen institutions or networks, the questions important to programme partners and other stakeholders come to the fore. They will often wish to know both how to strengthen their capacity and how increase involvement of target groups in the initiative. Appropriate approaches and methods are often participatory and collaborative.

Evaluations can also be designed with the intention of developing or shaping the initiative on an ongoing-basis. They can be similar to evaluations concerned with improving implementation, but there is an additional expectation that changes will be made to the nature of the evaluation object (the initiative) itself. Evaluation questions will, therefore, be related to questions of programme or project design. Methods of action research are traditionally associated with this purpose.

It can be helpful to think of evaluation criteria, or criteria for determining the success of the initiative, as being clustered around different types of evaluation purpose. These criteria can, in turn, be translated into clusters of specific evaluation questions. Criteria and questions that are relevant to an evaluation will vary from initiative to initiative and will depend on the results of the mapping exercises described above. As an illustration, examples of typical evaluation criteria and associated evaluation questions that might be applicable in the Science and Society domain are given in Table 5-2 below.

Table 5-2 Evaluation criteria and associated questions

Criterion type	Criterion	Associated evaluation questions
Structural / operational	Conceptual coherence	Did the theoretical base underpinning the project objectives 'hang together' and were the concepts plausible in terms of state of the art thinking?
	Accessibility	Did the institutional arrangements serve the needs of the target population or were certain groups excluded? Were all groups fully engaged in all aspects of the activity?
Content / outcomes / summative	Relevance	Were the materials produced appropriate for the target users?
	Participant satisfaction	Was the intervention received as positive or negative by participants? Were participants fully engaged in all aspects of the activity?
	Participant benefits	Are participants more aware of STEM educational and career options? Are participants more confident in expressing views about the role of Science and Society? Are teachers better able to communicate STEM concepts and enthuse their students about STEM?
	Stakeholder benefits	Have new collaborative networks been established? What are the barriers / drivers for the successful operation of collaborative networks? Do teachers feel more confident about their ability to carry out a hands-on STEM activity with their students?
	Cost-effectiveness	What is the ratio of costs to benefits for the range of projects funded as part of the programme / parent initiative?
Learning	Transferability (internal)	Can the lessons from this initiative be used to improve the programme in the future?
	Transferability (external)	Can the lessons from this initiative be used to develop or improve other programmes? What best practice has been identified? How should best practice be assessed?

5.5.3. Typical scoping and mapping design methods

The methods used during the initial phase of an evaluation typically comprise:

- ***Orientation activities*** which will offer an opportunity to review available information sources, clarify links with stakeholders and familiarise the evaluation team with the initiative.
- ***Model building studies*** which will seek to elicit underlying theories and models about Science and Society which inform the initiative, and clarify and adapt, if necessary, the measures and indicators used to conduct the impacts and process assessments. The mapping tool has been designed so that an examination of the distal dimensions listed will support this process. Data capture typically involves interviews with stakeholders and content analysis of relevant documentation.
- A preliminary description of the ***programme architecture*** of the initiative. It will include an audit of available documentation, covering policy, administration and management documents. This is normally carried out through interviews with project staff and content analysis of relevant documentation.
- ***A stakeholder analysis*** to map different perceptions of what stakeholders expect from the evaluation and identify potential differences and conflicts. It will also consult with relevant stakeholders to develop the infrastructure and logistics necessary for the delivery of relevant quantitative impact indicators.
- ***Initiative Audit.*** This will produce a concise description of the initiative on the ground, setting it in the context of its socio-economic and cultural environment; the scenario of delivery; and organisational structure. The mapping tool has been designed so that an examination of the proximal dimensions listed will support this process. Data capture methods would normally include: analysis of monitoring data; site visits; interviews with stakeholders.

Depending on the overall purposes of the initiative, a further set of activities might develop collaboration and review procedures and systems for the evaluation – for example, to engage the public involved in the initiative in the evaluation design and in subsequent evaluation monitoring and review, and to develop mechanisms for ongoing reflexive review of the evaluation and its outputs. This could include steering groups, citizens' juries and similar set-ups.

5.5.4. The importance of scoping and mapping design tasks

Given the nature of the domain, and its complexity, we would also propose that part of the initial preparatory work is devoted to providing consultancy time by a skilled evaluator to identify modes of enquiry and evaluation questions. This could be carried out within the context of a Science and Society Evaluation Advisory Facility to kick-start evaluation activity across the board within an initiative.

The amount of work that takes place *before* an evaluation design can be finalised is often surprising to new evaluation professionals and those unfamiliar with evaluation design. It is important that both the scoping and 1st stage design phase of an evaluation be adequately resourced, especially in terms of time. Best practice does *not* consist of trying to fit a preferred but potentially inappropriate evaluation method to any situation regardless of the nature of the initiative or the context within which the evaluation takes place.

Like all evaluators, those operating in the field of Science and Society need to resist the temptation to move directly to the selection of evaluation methods and techniques before a detailed assessment of the attributes of the initiative and the environment in which it operates has been completed. In this Evaluation Framework, we strongly argue that the selection of methods and techniques is a 2nd stage design task, along with the practical considerations of timescales and resources.

5.6. Stage 2: Method design tasks

Once the object and purposes of the evaluation have been established and the key evaluation questions and criteria specified, much of the hard work of evaluation design has been completed and it is time to move on to the stage 2 design task of assessing which research methods are most appropriate for answering the evaluation questions. To say that "once the stage 1 design tasks have been completed then the appropriate methods and associated tools and techniques[17] will 'drop out' of the equation" is to understate the work involved in this latter stage of evaluation design. However, most of the information needed to make an informed decision about which elements to select from the large range of available methods should now be in place. It should be noted that it will more often than not be appropriate to employ a number of complementary methods in a single evaluation. Indeed, this methodological triangulation can help to boost the internal validity of an evaluation.

The RCUK (2005) Practical Guidelines include a valuable introduction to the various research methods available to evaluators. In this Evaluation Framework we have assumed a degree of familiarity with research methods and their various strengths and weaknesses, although a summary of research methods and techniques, together with their relative pros and cons, is provided in Table 5-3. Given the varied nature of the evaluation questions that could be asked of Science and Society initiatives, it may be necessary for evaluation teams to include evaluators with a broad overview of available research methods as well as those specialised in working with mixed methods, quantitative methods and qualitative methods.

[17] Some definitions: Methodology – assumptions about what constitutes knowing (ontology), what constitutes knowledge (epistemology) and how this can be operationalised. Examples include phenomenology, constructivism, realism, and experimentalism; Method – a general mode of gathering data. Examples include interviews, RCTs, focus groups and surveys. To be distinguished from method of analysis. Technique refers to a specific means of effecting a particular method. Examples include self-assessment questionnaires, semi-structured interview, content analysis; Tool – a tool which supports the application of a techniques. Examples include topic guides and coding frames (ARTICULATE/ Tavistock Institute 1995).

Table 5-3 Methods and associated techniques

METHOD	Typical techniques	Typical context of use	Pros and Cons
SURVEYS	Interviews Mapping Questionnaires	All-purpose. Operational: mapping interactions between actors. Summative: user satisfaction; user impacts. Learning: surveys of participants experiences.	Easy to carry out. Can produce large numbers of responses. Limited depth in questionnaire surveys (less in interviews and focus groups). Good in outcome-linked evaluations.
FIELD STUDIES	Observation Task Analysis Critical incidents Ethnography Case studies Diaries	All-purpose. Summative: how users respond to intervention. Operational: how institutional structures operate. Learning: retrospective analysis of what happened. Comparison of different settings.	In-depth data, giving insights on social construction of intervention. Time-consuming and skill-intensive. Difficult to utilise in outcome-linked evaluations.
MODELLING	Simulations Soft systems	Usually operational and learning modes. Assessing organisational structure, dynamics and change. Cost-benefit analysis. Optimisation of management functions.	Can predict possible outcomes to adjustments in uncertain and complex contexts. Sometimes highly abstracted. Requires high level of skill.
INTERPRETATIVE	Content analysis	All purpose. Used in operational (analysis of meetings etc.); summative (analysis of materials or reports) and learning (deconstruction of programme reports).	Deconstruction of 'hidden' meanings and agendas. Rich interpretation of phenomena. Inherent risk of ideological bias.
CRITICAL	Discourse Analysis	More theoretical (usually critical theory) based than content analysis. Typically used to assess structure, coherence and value of large-scale programmes for 'learning' purposes.	As for interpretative methods, but emphasises establishment of generalisable laws. Perceived to be 'unscientific', especially by 'experimentalist' practitioners.
PARTICIPATORY	Action research	Typically in developmental evaluation mode.	Encourages real engagement of 'subjects' of intervention. Good in highly uncertain contexts. Evaluators sometimes get 'too involved' in intervention itself.

As argued above, the selection of methods and techniques will be highly dependent on the object and purposes of evaluation. In this respect, some broad ‘rules of thumb’ to consider are:

- ‘Operational’ evaluations that concentrate on providing real-time monitoring and support for project management typically entail the on-going collection of data from a limited number of key actors in the project (for example, project managers, materials designers and representative groups of target users). Commonly-used techniques are structured interviews; focus groups; diaries and logs; and task analysis (getting key actors to ‘think aloud’ about why they make a particular decision).
- Evaluations that focus on ‘accountability’ aspects such as the institutional arrangements of an initiative and how that initiative was managed typically involve: surveys of participants; critical incidents analysis (why certain decisions were taken and what were the consequences); socio-metric mapping (plotting interactions and communications patterns between key actors); systems modelling (what would have happened if one part of the structure had been changed); and content analysis (of meetings etc.).
- ‘Effects’ evaluations that consider, for example, the ways in which target users responded to a Science and Society intervention, and in what ways their behaviours changed, typically utilise: questionnaire surveys; interviews and focus groups; and critical incidents analysis. These methods are normally used retrospectively. In contrast, observation, diaries and logs, and content analysis (of, for example, effectiveness and participant satisfaction with regard to awareness raising events) are normally used in real-time as the intervention develops.

Another issue influencing the selection of particular methods and techniques is the epistemological orientation (or paradigm preference) of the main audience for evaluation outputs. In broad terms, evaluations focusing on success, outputs and outcomes will favour experimental methods using quantitative data. However, certain non-experimental techniques lend themselves to a more ‘quantitative’ orientation (and mode of analysis) than others. For example, large scale questionnaire surveys are frequently used in evaluations that have a summative (effects) component because they can provide (as a result of the application of sampling and probability techniques) estimates of generalisability in terms of impact to whole populations. Similarly, evaluators (and evaluation audiences) who are more interested in the different social constructions different cultural groups have with regard to, say, awareness raising campaigns may be more persuaded by the use of ethnographic techniques. At the other end of the paradigmatic ‘scale’, discourse analysis or other interpretative methods based on critical theory may be used to deconstruct how the ‘cultural logic’ of an initiative is constructed. It should also be remembered that the selection of particular methods and techniques has skill and data resource implications. Some techniques (for example, ethnographic methods) involve lengthy, in-depth field work producing copious amounts of data that require highly skilled analysis and interpretation.

We need to bear in mind, however, that our ‘ideal’ evaluation design may not be practical. An evaluability assessment, which assesses whether or not the conditions exist for a programme or project to be evaluated usefully within the time and budget allowed, may be worthwhile. Certainly, the availability of data necessary to the evaluation should be checked: it is not uncommon for evaluators to discover that secondary data sources upon which they rely are not robust enough or appropriate for their purposes or that access to certain sources cannot be negotiated.

5.7. Stage 3: Implementation and analysis

Having decided on which methods and techniques to use, important things to consider in the implementation and analysis stage of the evaluation include:

- ***Contingency planning:*** as with planning an evaluation in general, anticipating adjustments and changes to data collection is to be encouraged. It is useful to have a ‘plan B’ with alternative arrangements for data collection should it become apparent that, for example, time, skills or operational constraints are likely to conspire against planned activities.
- ***Triangulation:*** the evaluation should already have been designed with regard to the resource requirements of the choices specified and with the ‘insurance’ of contingency planning in mind. It is also worth noting that ‘insurance’ also has a methodological component: triangulation. Triangulation means utilising different methods to cover the evaluation from different angles (for example, assessing the effectiveness of organisational structures of a project from the points of view of different actors).
- ***Analysis requirements:*** it should be borne in mind that the selection of particular methods and techniques also implies using the appropriate type of data analysis (which has its own resource and skills implications). In general, large data sets (such as derived from surveys) normally need statistical software systems such as SPSS. Interpretative data (derived, for example, from content analysis) can be analysed with proprietary qualitative software packages such as NVivo. In any case, a clear coding frame to analyse such data is necessary.
- ***Operational rules:*** the evaluation should be able to track (and have a record of): what data are being collected, who collects the data, and in what form and location the data are stored. Clear rules about operational procedures should be set out and distributed to all those involved in data collection and analysis. Similarly, it is useful to draw up ‘evaluation contracts’ with other stakeholders, especially those supplying information. These contracts should specify the objectives of the evaluation and any guarantees that apply (for example, on confidentiality).

5.8. Stage 4: Reporting and Dissemination

This phase should be a continuation of the evaluation process. In this sense, it is important to give the initiative being evaluated, as well as the evaluation itself, and project participants a sense of 'closure' of the project and the evaluation, where appropriate, by running concluding feedback events.

More generally, it is important to the reputation, value and impact of the evaluation to give final formal feedback to everybody who has contributed in some way to the evaluation (by sending them a copy of the report or inviting them to a final feedback event).

Dissemination should not be restricted to the circulation of a final report – especially in the case of 'developmental' process evaluation. Different stakeholders may require different communication approaches. These might include:

- Short summaries of the evaluation, tailored to different audiences
- Journal articles for other researchers
- Topical articles in the 'trade' press
- Workshops for specific audiences
- Feedback seminars for key decision-makers.

The use that is made of dissemination outputs should be consistent with the 'purposes' of the evaluation as defined in the initial Preparatory and Design phases. In other words, evaluations should be designed in terms of the decisions and actions they will inform. It is not always easy to reflect this in recommendations, especially when the relevance of such recommendations may not be easily recognised by some stakeholders. The art of making useful recommendations lies in:

- Understanding the context in which the audience operates.
- Addressing future realities rather than dwelling on the past.
- Clarifying choices based on realistic options.
- Showing how recommendations can be implemented in practice.

Part III: Applying the Evaluation Framework

6. MAKING THE EVALUATION FRAMEWORK WORK IN PRACTICE

6.1. What this section is about

In this section we apply the evaluation principles and procedures set out in section 5 to the particular evaluation issues and challenges posed by Science and Society programmes and initiatives. Its starting point is how to define the ‘object’ of evaluation: essentially the ‘thing’ that needs to be evaluated. The section demonstrates how defining the ‘object’ of evaluation shapes the selection of appropriate evaluation approaches and methods.

6.2. Introduction

We have tried to stress in the Framework the importance of avoiding falling into the ‘contingency trap’. The trap is to assume that mapping the characteristics of a Science and Society initiative in isolation will automatically reveal the correct evaluation recipe and the ingredients – the evaluation methods and techniques – required to make up the recipe. It has been argued instead that effective evaluation draws on understandings about elements such as the environment in which the initiative operates; the stakeholders involved; the purpose of the evaluation; and the kinds of questions that need to be asked. These understandings need to make sense of the complexity and ‘messiness’ of the Science and Society domain and the way it reflects a wide range of stakeholder views and different types of knowledge and beliefs. Selection of appropriate methods and techniques falls out of these understandings, rather than simply an understanding of the properties of the initiative being evaluated.

Evaluation design – and implementation – is an iterative process in which a picture is built up about what is being evaluated; why it is being evaluated; who the evaluation is for; and how the evaluation should be carried out. These elements feed into each other to shape the evaluation design. Yet this poses a number of dilemmas for those commissioning evaluation, as well as for evaluators themselves – what is an appropriate starting point for this iterative process? How can complexity and messiness be managed?

Evaluation in this context is a bit like detective work. The evaluator needs to look for clues that can help build the evaluation picture and can help to signpost a path through the complexity. To identify the clues, the first thing to do is: define the ‘object of evaluation’. Put simply, the ‘object’ of evaluation is the thing that needs to be evaluated. This ‘object’ reflects a number of attributes that distinguish it from other ‘objects’ and which provide the clues and signposting to help construct an effective evaluation design. Mapping the ‘object’ of evaluation in relation to a set of attributes helps the evaluator get a sense of the kind of evaluation approach that is likely to be suitable. For example, there is obviously little point in designing an evaluation that incorporates a large scale longitudinal survey for a one-off science road show outside the Science Museum. This in turn will give some clues as to the appropriate methods and techniques that could be used to collect and analyse data. It is important to stress once more that this does not imply a 'contingency' model. Rather, the evaluator has to make a 'heuristic', or 'rule-of-thumb, judgement based on the assembled 'clues' about the best evaluation approach to use.

This section discusses ways of applying the Framework to develop effective evaluation strategies, using the 'object of evaluation' as a starting point. It:

- Provides a tool to help define the 'object of evaluation'.
- Links defining the object of evaluation to selection of appropriate evaluation approaches.
- Provides a set of examples of Science and Society initiatives to illustrate how evaluation choices might be made.
- Provides an elaborated 'case study' to illustrate in detail how a particular evaluation approach and methodology works.

6.3. Defining the object of evaluation

Our study has highlighted a range of 'attributes' that could be used to define the 'object' of evaluation in the context of Science and Society. These include:

- ***Type*** – which defines in broad terms the 'essence' of the initiative with regard to what it sets out to do.
- ***Purposes and objectives*** – which define what is expected to be produced and achieved.
- ***Properties*** – its distinguishing features, in terms of, for example, scale, time and complexity.
- ***People*** – the target groups and audiences aimed at; the stakeholders involved.
- ***Delivery scenario*** – the actual delivery mechanisms involved in implementing the initiative.
- ***Innovation level*** – the extent to which the initiative is trying out new and experimental things or using tried and tested approaches.

Again though, we stress that, while the project work has suggested that this is a helpful 'way in' to think about evaluation design in practice, it is just that – a 'way in' – and not an alternative to considering the complexity of Science and Society initiatives and the environment within which they operate.

6.4. Defining the broad evaluation approaches

Similarly, the study has suggested that there are five broad approaches or combinations of approaches that could typically be used in evaluating Science and Society actions and initiatives:

- ***'Effects' approach.*** This has a 'summative' purpose and is intended to establish what works, for whom, under what conditions.
- ***'Planning' approach.*** This aims to support the scoping and design of the initiative prior to its implementation.

- ***'Developmental' approach.*** This aims to support the evolution of the initiative through, for example, improving its capacity as it develops.
- ***'Operational' approach.*** This aims to review and support the implementation of the initiative by focusing on its processes, for example, management structures.
- '***Accountability' approach.*** This aims to assess the degree to which and in what ways resources have been utilised by the initiative.

Most evaluations that are commissioned are of the 'effects' type. It is useful to distinguish between three different kinds of 'effects':

- ***Outputs*** – the things an initiative produces (for example STEM teaching resources).
- ***Outcomes*** – the immediate effects associated with the use of an initiative's outputs (for example, the effects of a 'teacher fellowship' scheme on science teachers' skills).
- ***Impacts*** – the longer term effects of the initiative's outcomes (for example, the effects of physics teachers' improved skills on their pupils' achievements and career choices).

These three types of effects are often confused and are often mixed together. Clearly, to a large degree, the three types equate to the timescales of an initiative. Within the context of Science and Society, 'impacts' frequently imply changes in the behaviour of 'end users', such as citizens, teachers and students. These changes take place over a relatively long period of time, and evaluations of initiatives seeking to effect such changes should ideally be conducted over a long period of time. Similarly, 'outcomes' usually imply shorter-term changes, for example, changes in attitudes associated with a public awareness campaign. An evaluation of the outcomes of a 'one-off' event, like a road show, would normally be limited to eliciting immediate feedback from the audience about whether and in what ways their perceptions of science had changed. Longer term evaluation methods would be required to determine if the impact of the event was lasting.

However, the exact nature of the expected outcomes and impacts of an initiative also depend on the type of initiative being implemented. For example, a programme with the objective of developing the funding infrastructure necessary to support local partnerships in order to create and deliver STEM initiatives may not have a 'direct' impact on changing school students' career aspirations, but may have significant impacts on the national STEM skills base further on in time. This brings into play another feature of Science and Society initiatives that must be taken into account in designing evaluations: the position they occupy in the 'delivery chain'. Broadly speaking, it is difficult to evaluate the impacts of an initiative on 'end user' behaviours the further removed it is from directly engaging with these users. This is mainly because of the effects of 'intervening variables'. These issues are discussed in more detail below.

Linking the object of evaluation with evaluation approaches

As a general rule, the specific attributes of a particular 'object' of evaluation will provide clues as to the type of evaluation approach that could usefully be deployed.

In turn, the different evaluation approaches are themselves broadly associated with particular evaluation questions, measurement criteria and methods and techniques. The following sections illustrate how this works.

6.4.1. Type of initiative

Table 6-1 summarises the broad types of initiatives identified by the study, together with the evaluation approaches, questions, criteria and methods that would typically be associated with a particular type of initiative.

Table 6-1 Science and Society initiative types and evaluation approaches and methods

Type of initiative	Evaluation approach	Typical evaluation questions and criteria	Typical evaluation methods
Awareness-raising campaigns	Effects: outputs; outcomes; impacts Accountability	What kinds of people changed their attitudes towards science and in what ways?	Cross-sectional surveys Focus groups Content analysis of media
Public participation	Planning Developmental	How can public anxieties about nuclear power be productively harnessed to develop sustainable energy?	Citizens' juries Focus groups Action research
Interactive events (outreach; theatre; demonstrations)	Effects: outputs; feedback	How many and what type of people attended the event? How engaged was the audience? In what ways did participants' views of science change?	Exit polls Quota sample Analysis of attendance records Observation Interviews
Education and training	Effects: outputs; outcomes; impacts Operational Accountability	The number of high school students completing science courses Movement in the salary levels of scientists and technologists	Statistical analysis Questionnaire surveys Interviews
Ongoing profile-raising	Effects: outputs; outcomes; impacts	To what degree and in what ways is science covered in the popular media? What contribution does profile-raising investment have to science policy and improving the knowledge base?	Content analysis of sample of newspapers Citation analysis of academic journals
Targeted access and inclusion actions	Effects Operational Accountability	Have the proportions of black and ethnic minority students achieving science degrees increased?	Statistical analysis Questionnaire surveys
Policy actions	Effects: outputs; outcomes; impacts	Has the implementation of the consultation exercise created new partnerships?	Focus groups Documentation analysis
Horizontal and supporting actions	Effects: outputs; outcomes Operational Accountability	How many schools have taken advantage of subsidies for Science Learning Centres?	Statistical surveys Documentation analysis
Operational Reviews	Effects: outputs; outcomes Accountability	Which public engagement approach is most cost-effective?	Process evaluation Cost-effectiveness analysis

6.4.2. Scale

The scale of an action, initiative or intervention can be measured in very simple terms: small, medium or large, and gradations of these three levels (very small, very large and so on). In broad terms, therefore, they are typically distinguished by:

- Single events – such as a local 'science fair'.
- Multiple events (small scale) – such as a series of science fairs.
- Projects (medium scale) – of more extensive scope and objectives than 'events'.
- Programmes (large scale) – incorporating combinations of events and / or projects, with multiple stakeholders and multiple objectives.

With single event initiatives, evaluators are typically interested in 'effects', but outcomes and impacts are difficult to assess. In turn, such initiatives do not require much investment in management and administration, so formal operational evaluation is hardly ever necessary. However, scale also reflects dimensions such as: the spatial area covered by the initiative (local; regional; national; trans-national); the cost of the initiative; or more ephemeral factors, such as the comprehensiveness or magnitude of the initiative's objectives. These factors in turn reflect the 'properties' of the initiative. For example, policy actions may encapsulate a 'grand vision' for Science and Society, incorporating sweeping and radical changes to existing policy provision on a grand scale or may target a limited and particular segment of the existing provision. Public awareness campaigns may target the whole population or targeted segments. They may reflect an extensive spectrum of Science and Society issues or may focus on a single issue.

The scale of an evaluation can reflect the scale of the initiative in an absolute or relative sense. An evaluation of a small scale initiative implies a small scale evaluation in the absolute sense. An evaluation of a small segment of a programme may, however, entail a relatively small scale evaluation.

Given this intrinsic variability of the domain, defining 'scale' from the point of view of an evaluation requires a heuristic judgement by the evaluator, more or less involving a classification based on our starting position above – small, medium or large.

Table 6-2 summarises the evaluation approaches, questions, criteria and methods that would typically be associated with different scales of initiative.

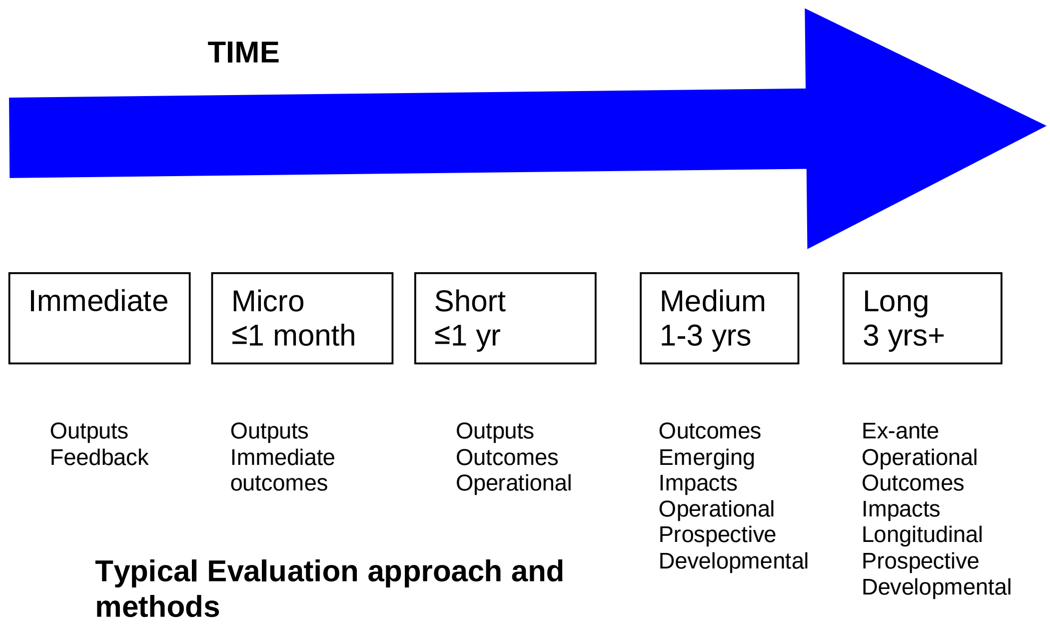
Table 6-2 Typical scales of Science and Society initiatives and evaluation approaches and methods

Scale of initiative	Evaluation approach	Typical evaluation questions and criteria	Typical evaluation methods
Single events	Effects: Outputs; immediate outcomes	How many and what type of people attended the event? How engaged was the audience? In what ways did participants' views of science change	Exit polls Quota sample Analysis of attendance records Observation interviews Self-evaluation
Small	Effects: Outputs and outcomes Operational	How many and what type of people attended the event? How engaged was the audience? In what ways did participants' views of science change? How effective was the organisation between the events?	Exit polls Quota sample Analysis of attendance records Observation Interviews Process Review
Medium	Effects: Outputs and outcomes Operational Accountability Developmental	In what ways did participants' views of science change? How effective was the project management and administration? In what ways could the project be supported as it evolves? Was the project value for money?	Statistical analysis Questionnaire surveys Interviews Focus groups Documentation analysis
Large	Effects: Outputs, outcomes, impacts Operational Accountability Developmental Planning	The number of high school students completing science courses Movement in the salary levels of scientists and technologists	Ex-ante evaluation Action research Statistical analysis Longitudinal studies Interviews Case studies of projects Process review Cost-benefit analysis

6.4.3. Timescale

As discussed above, the different types of effects that an initiative is designed to bring about are associated with different timescales. In turn, the timescale of an evaluation will determine how and in what ways those effects can be measured.

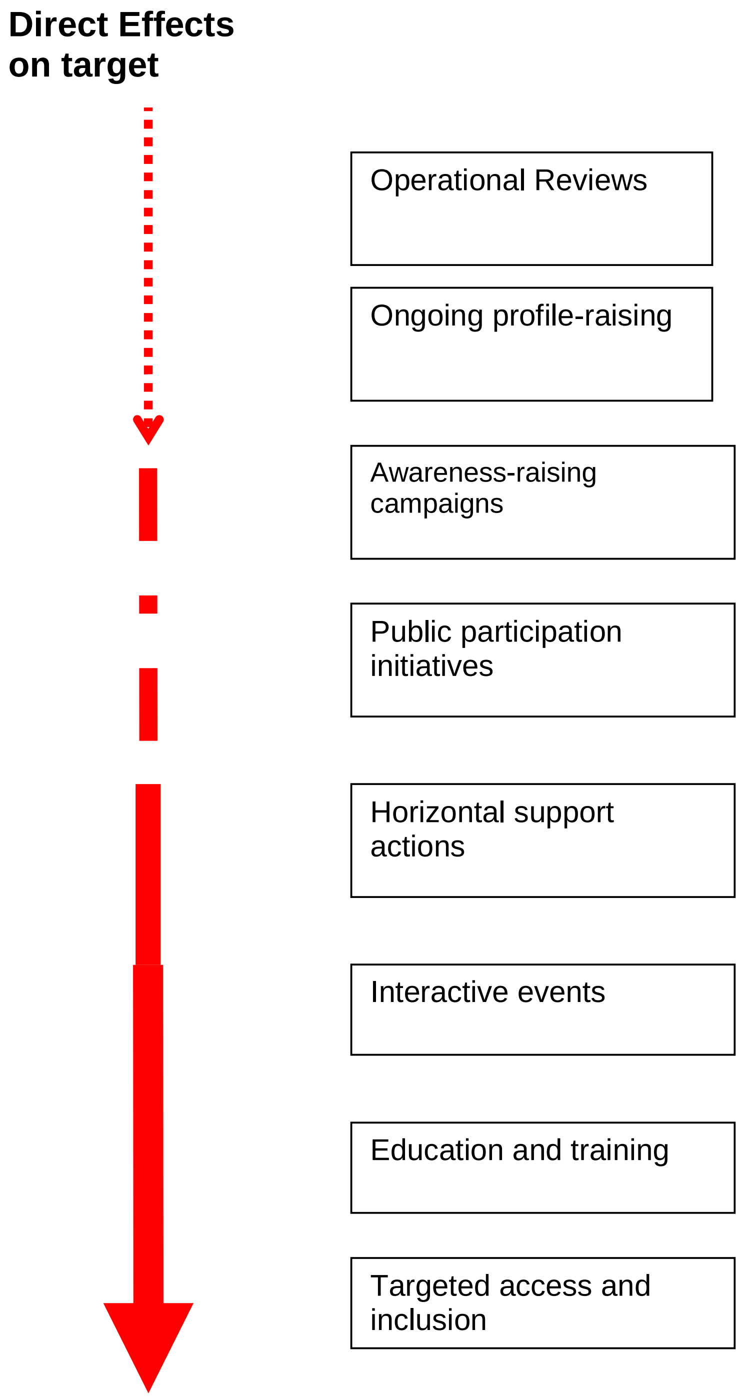
Figure 6:1 Relationship between initiative timescale and evaluation approach



6.4.4. Delivery chains and distance from beneficiaries

As discussed in section 4.6 above, understanding the nature of the delivery chain of an initiative is helpful in evaluating its impact. We revised significantly our conception of the shape of delivery chains in the Science and Society domain as a result of the SETNET pilot. The length of an initiative's delivery chain, or the 'distance' of the initiative to its beneficiaries will differ depending on what type of initiative is being considered. Figure 6.2 illustrates the distance between some of the main types of Science and Society initiatives and their beneficiaries.

Figure 6:2 Typical distance of different types of Science and Society from beneficiaries



7. TYPICAL SCIENCE AND SOCIETY EVALUATION SCENARIOS

7.1. What this section is about

This section takes the principles, procedures and guidelines set out in Section 6 and illustrates how they can be applied in practice with reference to a set of examples of a range of Science and Society initiatives. The examples have been selected to reflect a broad spectrum of the range of initiative types, scales, timeframes and objectives identified in our review of the domain, as discussed in Part 1 of this Framework. Each example is illustrated with reference to the kind of evaluation strategy that could typically be applied for that particular case. The evaluation strategy illustrated provides guidance on overall evaluation approach; stakeholder analysis; definition of the 'object' of evaluation; evaluation purposes; questions and criteria; selection of methods and techniques. The main purpose of this section is to help evaluators choose the most appropriate strategy, approach and design to suit a particular situation.

7.2. Introduction

As discussed above, the Framework developed in this study specifically avoids adopting a contingency-based approach to evaluation design. For reasons discussed above including: the complexity of the Science and Society domain; the rapidly evolving and contested nature of scientific knowledge; the wide variety of different initiatives and activities, and their correspondingly highly differentiated characteristics, it is virtually impossible to configure an 'exact' combination of evaluation approach, methods and techniques that corresponds to a particular set of Science and Society circumstances and characteristics.

Nevertheless, as outlined above certain kinds of Science and Society initiatives lend themselves to a particular kind of evaluation. In this section we provide practical guidelines on choosing the most appropriate and effective evaluation approaches, methods and techniques. This is not based on a 'contingency' approach. Rather, we adopt more of a 'benchmarking' approach that allows those wishing to commission or carry out an evaluation to compare their own situation with a range of typical evaluation situations and scenarios, and with regard to a range of exemplars of Science and Society initiatives. The examples have been selected to reflect a broad spectrum of the range of initiative types, scales, timeframes and objectives discussed above. These exemplars are 'hybrid' cases, incorporating elements of real initiatives and hypothetical ones. They are summarised in the table below.

Table 7-1 Summary of Evaluation Examples

Example	Description	Type	Scale/Duration
Plant Takeaway	Exhibition to raise awareness of biology and environmental issues	Interactive event	Micro. Local one-week event
Lecture Master	On-line information service on scientific lectures	Ongoing profile-raising	Restricted audience. Ongoing

Example	Description	Type	Scale/Duration
Small Grants initiative	Scheme to encourage schools to apply for small grants to promote STEM teaching	Support Action	National. 3 year duration
Dialogue	Developing a collaborative learning model and methodology to engage public in developing policy	Participation	National. Short-term (6 months)
Acrisat Plus	Educational initiative to encourage black and ethnic minority pupils to study STEM and consider STEM careers	Educational	National. Medium-term (3 years)
Prometheus	National programme to promote innovation in public awareness and STEM education	Policy action Awareness-raising. Support action Education	Very large. National. 5 years

7.3. EXAMPLE 1: Plant Takeaway

Description:

Plant Takeaway is an exhibition running for one week only, housed in a temporary space within a UK Science Exhibition Centre, and supported by the National Endowment for Science Technology and the Arts (NESTA). It aims to communicate human-plant interdependency by building a life-size automaton of a kitchen scene. The exhibition is aimed at the general public and its main objective is to increase awareness of science issues around biology, the environment and sustainability by situating these issues within an everyday context that ordinary people can readily identify with.

Main objective:
To raise citizens' awareness about biology, the environment and sustainability.
Type:
Interactive event.
Duration:
Micro (1 week).
Delivery:
Interactive exhibition on single site.
Scale:
Small. Target 1,000 visitors.
Target groups:
General Public.
Outputs:
Interactive multimedia learning tool. Information leaflets.
Expected outcomes:
1,000 visitors engage with exhibit.
Expected impacts:
Improved public understanding of plant biology processes. Increased awareness of environmental issues.
Innovation and complexity:
Innovation limited to pedagogic approach and level of interactivity.
Position in delivery chain:
Downstream. Engages with end users and aims to change attitudes.

Evaluation Approach:
The main factor shaping the evaluation in this case is the very short duration and small scale of the initiative. This suggests that the evaluation needs to focus on the outputs of the initiative and their effectiveness in delivering its key objective: to engage visitors in thinking about plant biology and related broader issues such as the environment. The short time frame of the initiative, together with the scenario in which the exhibition takes place, limits the extent to which outcomes can be assessed. The ‘random’ nature of the participants’ engagement with the exhibition would make it difficult to carry out a more classic ‘experimental’ approach, involving a randomly selected ‘control’ group of a sample of the audience and a ‘comparison group’ of people who had not visited the exhibition, and assessing whether any differences in attitudes to environmental issues between the two groups could be determined. The costs of implementing this type of evaluation approach are also likely to be prohibitive, especially in comparison with the relatively small costs of the exhibition itself. A typical evaluation strategy in this case would be to focus on eliciting feedback from visitors through an ‘exit poll’ using a short attitude questionnaire, including rating scales on items such as user satisfaction.

Table 7-2 Plant Takeaway

Evaluation Object	The automaton and its effectiveness as an awareness-raising tool
Evaluation Purposes	***Effects*** Evaluate the outputs of the project, focusing on the automaton and supporting literature. Assess the extent to which and in what ways the exhibition has influenced the audience’s views on plant biology and environmental issues. ***Accountability*** Assess the cost-effectiveness of the scheme.
Key Stakeholders	NESTA Exhibition Centre Managers
Evaluation Questions	***Effects*** How many visitors attended the exhibition? How effective was the exhibition in engaging and maintaining the interest of visitors and in what ways? In what ways did the exhibition affect awareness of and attitudes towards plant biology and environmental issues? ***Accountability*** Does the scheme provide good value for money?
Evaluation Criteria	***Effects*** Attendance rates Usability and user satisfaction of automaton. Self-reported increase in audience awareness. ***Accountability*** Exhibition set-up and operational costs per participant.
Methods	***Effects*** Audience head count. Baseline analysis with attendances at comparable exhibitions. Expert review of exhibition design and learning model. Exit poll – self administered questionnaire using audience quota sample. ***Accountability*** Value for money analysis. Comparison with similar initiatives.

7.4. EXAMPLE 2: LectureMaster

Description:

LectureMaster is an on-line list of all the lectures with a 'scientific' focus that are taking place throughout the UK within a rolling timeframe of one year. The initiative provides a diary to enable people interested in the subject matter to check out what is taking place and, in some circumstances, book a place at a particular lecture. The system also alerts its members, whose details are stored in its database, to lectures with content that fits with their profiles and interests. Users can evaluate and rate the lectures, and the scientists who deliver them, against a range of evaluation criteria. They can take part in on-line seminars on particular scientific themes.

Main objective:
To distribute information about scientific events and circulate scientific knowledge.
Type:
Ongoing profile-raising.
Duration:
Ongoing.
Delivery:
Website.
Scale:
Small. Potentially international audience but restricted audience base.
Target groups:
General Public.
Outputs:
Interactive website. Distributed knowledge base.
Expected outcomes:
20,000 visitors per month.
Expected impacts:
Increase attendance at lectures and improve delivery of information to those likely to attend.
Innovation and complexity:
Innovation limited – basic website functionalities – but partnership arrangements novel.
Position in delivery chain:
Upstream. Main objective is to maintain position of science in public discourse.

Evaluation Approach:
This initiative is primarily intended to maintain the profile of scientific discourse by exposing the public to scientific developments and scientists, and keeping scientific issues at the forefront of debate. The main evaluation focus is, therefore, to assess the extent to which it contributes to supporting the scientific knowledge base. This calls for evaluation methods that are normally used in academic publishing – the use of citation analysis. Another key evaluation issue is to assess the effectiveness of its innovative partnership and organisational structure, and their transferability to other situations. The evaluation is also concerned with exploring in what ways access to the service shapes users' understanding of scientific issues and leads to changes in behaviour – for example, whether exposure to the service leads to attendance at lectures and whether this in turn supports other activities, for example, learning and employment strategies. The timeframe of the initiative allows scope for implementing a limited longitudinal study to explore the behaviours of a sample of the service user population. The technological platform also offers opportunities to use more

innovative evaluation techniques – including on-line questionnaires and content analysis of on-line seminars.

Table 7-3 LectureMaster

Evaluation Object	The delivery vehicle – including the website and database together with the organisational arrangement in which it operates.
Evaluation Purposes	***Effects*** Evaluate the outputs of the project, focusing on the website and its content. Assess the utilisation patterns of the service. Assess the extent to which and in what ways the initiative is contributing to the development and implementation of policy and practice. Explore whether and in what ways the service contributes to changes in user attitudes and behaviours. ***Operational*** Evaluate the efficiency and effectiveness of delivery of the service. ***Accountability*** Assess the cost-effectiveness of the service.
Key Stakeholders	National Newspaper. Venture Capital company. Service managers. The scientific community. Service users.
Evaluation Questions	***Effects*** What is the level of use of the site? What is the service user profile? What do users gain from the service? Does use of the service lead to other behaviours, including attending lectures, learning more about science, further study and changes in employment? ***Operational*** How effective is the management and administration of the service? ***Accountability*** Is the service value for money?
Evaluation Criteria	***Effects*** Utilisation rates. Usability and user satisfaction of website. Ratio of site utilisation to attendance at lectures. Level and nature of citations of site. Self-reported effects on user behaviours. ***Operational*** Technical efficiency of website. Effectiveness of user support. Effectiveness of service management and administration. ***Accountability*** Service set up and operational costs per user.
Methods	***Effects*** Site citation analysis. Website log analysis. Content analysis of on-line seminars. On-line User Questionnaire. Longitudinal user survey. ***Operational*** Website technical log analysis. On-line User questionnaire. Stakeholder interviews.

	Site management interviews. ***Accountability*** Stakeholder interviews. Value for money analysis. Comparison with similar initiatives.

7.5. EXAMPLE 3: Small Grants Initiative

Description:

This initiative aims to support the transition of Scottish students from primary to secondary science education and to support teachers in making productive links between schools. Ultimately, it aims to stimulate children's enthusiasm for science and improve their attainment by enhancing their science experience. The project does not support direct STEM learning. It provides funding to encourage schools to apply for 'Science Small Grants'. This encouragement is delivered through workshops, one-to-one advice and consultancy, supported by a website which provides advice, downloadable application forms and a feedback questionnaire.

Main objective:
To boost science skills and knowledge in schools by encouraging teachers to apply for grants.
Type:
Support action.
Duration:
Medium (3 years).
Delivery:
Website; workshops.
Scale:
Medium – national target group of schools. 360 schools. 110 projects.
Target groups:
Science teachers.
Outputs:
Workshop; website.
Expected outcomes:
Grants awarded to 100 projects.
Expected impacts:
Increase investment available in schools receiving grants through matched funding. Improve knowledge transfer and collaboration between schools. Increase awareness of pupils around science.
Innovation and complexity:
Low. Scheme provides information and support, rather than novel experiments.
Position in delivery chain:
Middle range. Bridges policy initiative (Small Grants Scheme) and schools.

Evaluation Approach:
This scheme is essentially a support action. Its essential objective is to provide a bridge between a policy initiative – a 'Small Grants Scheme' to support STEM projects in schools – and the intended target group of the Grants Scheme – schools. It does this by spending resources on promoting the scheme to encourage schools to apply for grants. It is, therefore, positioned towards the middle of the 'delivery chain'. It has no direct engagement with teachers or pupils. The main purposes of the evaluation are to assess the effectiveness of the scheme in reaching this target audience and, subsequently, to enable teachers to successfully win funding for projects aimed at improving the effectiveness of STEM teaching in their school.

Table 7-4 Small Grants initiative

Evaluation Object	The scheme as a support action. How it bridges a policy initiative with schools.
Evaluation Purposes	***Effects:*** Evaluate the outputs of the project, focusing on its website and workshops. Assess the level and profile of schools applying for grants. Assess level and profile of applicants awarded grants. Examine the types of projects securing funding. ***Operational:*** Assess the extent to which the management and administrative systems of the scheme are 'fit for purpose'. Identify user satisfaction. ***Accountability:*** Assess cost-effectiveness of the scheme.
Key Stakeholders	Scottish Executive. Initiative managers. Science Teachers.
Evaluation Questions	***Effects:*** What has been the level of uptake of grants? What kinds of schools have participated in the scheme? What kinds of teachers have benefited from the scheme? What kinds of professional development have been supported by the grants? What have been the benefits for participating teachers? ***Operational*** How efficient and effective have the workshops and website been in supporting the scheme? How efficient and effective are the management and administration systems and procedures? ***Accountability*** Is the scheme value for money?
Evaluation Criteria	***Effects*** Workshop attendances. Website utilisation rates. Technical effectiveness of website. User satisfaction of website and workshops. ***Operational*** Response times for teacher enquiries. Level and type of problems experienced by stakeholders. ***Accountability*** Scheme operational costs per participant.
Methods	***Effects*** Statistical analysis of utilisation rates. Website logfile analysis. Teacher self-administered questionnaire survey. Observation of workshops. Focus groups with teachers. In-depth interviews with teachers. ***Operational*** Content analysis of management meetings. Interviews with key stakeholders. ***Accountability:*** Value for money analysis. Comparison with similar initiatives.

7.6. EXAMPLE 4: Dialogue

Description:

This initiative explores and applies ways of engaging the public in policy design. Its 'dialogue' theme incorporates four integrated actions. Firstly, a 'mapping' exercise to review state of the art in public engagement methods and practices. Secondly, utilising the results of the mapping exercise, an 'experimental' action aimed at exploring innovative ways of developing reflective and reflexive 'collaborative learning' environments, including 'action learning sets', to engage the public and other key stakeholders in policy dialogue. Thirdly, an implementation action, involving applying the discursive model developed in the previous phase to a consultation exercise aimed at establishing public attitudes to issues around 'trustworthiness' of Information and Communication Technologies. Finally, the development of a set of 'Good Practice Guidelines' to support the design, development and utilisation of ICTs. The initiative involves a partnership comprising government departments; a panel of academics and experts; representatives of industry and societies (including Wellcome Trust; British Association; telecommunications and software industry representatives); and citizens' panels (targeting 350 citizens in total).

Main objective:
To develop, test and apply an innovative participatory methodology to support public engagement and collaboration in science policy and practice.
Type:
Public Participation.
Duration:
Short (1 year).
Delivery:
Reviews and meta-analysis; workshops; citizens' panels; action learning sets.
Scale:
Medium – national in scope.
Target groups:
Citizens.
Outputs:
Model; methodology; Good Practice Guidelines.
Expected outcomes:
Improve understandings of engaging citizens in policy design; develop effective and transferable models to promote public engagement.
Expected impacts:
Improve effectiveness and relevance of ICT policy.
Innovation and complexity:
High. Experimental design and testing of new approaches to public engagement. Complex partnership and logistical arrangements.
Position in delivery chain:
Middle range.

Evaluation Approach:
This initiative calls for a developmental evaluation approach. Its focus is on promoting dialogue and collaboration through active participation with citizens. It is exploratory and experimental in nature and works in an area that is under-developed and where there is little evidence base. The adoption of a collaborative learning approach using action learning sets allows opportunities for the evaluation to work with the stakeholders through an action research approach. This entails working with stakeholders from the outset and throughout the duration of the initiative. It will

provide inputs to the evolving design of the participatory model and methodology. The evaluation design should also include a summative element, based on assessing the effectiveness of the outputs of the project – including the workshops and action learning sets – and covering the model and methodology. Finally, the main output of the initiative – the Good Practice Guidelines – would need to be evaluated against the project's objectives.

Table 7-5 Dialogue

Evaluation Object	The participative model and methodology, together with associated outputs including the Guidelines.
Evaluation Purposes	***Effects*** Assess the appropriateness of the mapping methodology used in the review, together with its results. Review the appropriateness and effectiveness of the collaborative learning model and methodology. Evaluate the outputs of the project, focusing on the collaborative learning methods, the participative methodology and Guidelines. ***Developmental*** Provide inputs to the initial design of the initiative and its methodology. Provide advice and support in setting up and implementing the action learning sets. Review the results of the action learning sets and advise on the selection and design of the pilot approach. Monitor and support the piloting phase of the initiative and evaluate its results. Assess the transferability of the model to similar policy areas.
Key Stakeholders	Department of Trade and Industry. Academic Steering Group. Scientific Societies. Commercial Partners. Citizens' Groups.
Evaluation Questions	***Effects*** How appropriate and effective is the mapping approach? How representative of state of the art are the results of the mapping activity? Are the citizens targeted and engaged in the initiative representative of the national profile? Are certain groups not included in the initiative? How relevant and effective is the model used in the action learning sets? How relevant and effective is the piloting methodology? How useful are the model, methodology and Guidelines? Are they transferable to similar contexts?
Evaluation Criteria	***Effects*** Comprehensiveness and relevance of the mapping review. Conceptual coherence and relevance of the collaborative learning approach. Representativeness of citizens' panels. Efficiency and effectiveness of participation actions. Efficiency and effectiveness of piloting actions. Technical coherence, effectiveness and usability of the Guidelines. Transferability of initiative outputs .
Methods	***Effects*** Review of state of the art. 'Goodness of fit' between state of the art and models used. Comparison of profiles of participants against baseline

	indicators. Content and discourse analysis of models, methodologies and outputs. Expert Panel review of models, methodologies and outputs. Observation of action learning sets. Content and discourse analysis of outcomes of action learning sets. Stakeholder interviews. ***Developmental*** Stakeholder mapping and needs analysis. Action research. Feedback and Reflection workshops.

7.7. EXAMPLE 5: Acrisat Plus

Description:

This initiative addresses the concerns raised in recent policy documents and initiatives over the low level of participation of black and ethnic minority groups in taking STEM subjects in education and their low representation in the STEM workforce. The initiative aims to help reverse these trends through a range of integrated activities. These include a consultation exercise with community groups and black businesses; research and workshops on the issues; raising awareness and encouraging organisations to tackle the issues in their area or sector; funding ‘exemplar projects’ to try out new approaches and ideas; and developing and disseminating a database of resources including good practice. A major element of this involves ‘outreach’ work, incorporating the use of a consultant to visit and liaise with black business organisations; school visits; residential workshops; and recruiting ‘role models’.

Main objective:
To improve the take-up and the perception of science and technology in education and at work among the UK residents of African Caribbean origin, and to advance the educational achievements and career aspirations of black youth within the fields of science, mathematics and technology.
Type:
Awareness raising; access educational.
Duration:
Medium (3 years).
Delivery:
Outreach (awareness-raising; neighbourhood events); targeted projects; website and database.
Scale:
National – but specifically targeted at black and ethnic minority communities.
Target groups:
Black and Ethnic Minority Organisations.
School students from black and ethnic minority groups.
Expected outcomes:
Raised awareness in the science and education sectors about barriers and good practice. Increase the number of organisations in the sectors undertaking initiatives to change under-representation and aspirations of African Caribbean people in the sciences. Joint working between the BME community and the science community.
Expected impacts:

Increase in the numbers of African Caribbean people taking up qualifications and careers in the sectors.
Innovation and complexity:
Complex partnership arrangements.
Position in Delivery chain:
Downstream. Direct engagement.

Evaluation Approach:
A distinguishing feature of this initiative is its capacity to engage with 'excluded' and 'hard to reach' target groups. This puts an emphasis on evaluating the effectiveness and efficacy of the underlying models and strategies of empowerment, the pedagogic approaches used and the assumptions the initiative makes about the barriers to participation in learning and employment, as well as the implementation approaches – including outreach work. This essential feature of the initiative also implies an emphasis in the evaluation on actively engaging target groups and other stakeholders in the design and implementation of the evaluation itself – using, for example, 'reference groups' to review how the evaluation is shaped and how it progresses. Against this background, the main challenge for the evaluation is to demonstrate what works, for whom and under what circumstances. The timescale of the evaluation will allow evidence to be collected on the medium-term outcomes of the initiative, focusing on awareness-raising and attitude change, but impacts – in terms of career decision. for example – will need to be confined to assessing potential impacts, such as expectations of STEM careers. The timescale does allow scope for carrying out a 'quasi-experimental' design. This could entail selecting a 'treatment group' of schools participating in the initiative, and a 'comparison group' of schools with a similar profile who are not involved in the initiative. The effects of participation can be measured in terms of indicators such as numbers deciding to study STEM subjects; school performance; self-reported career strategies; and science literacy. The website allows opportunities for data collection involving on-line surveys and discussion groups.

Table 7-6 Acrisat Plus

Evaluation Object	Whether and in what ways the initiative can potentially effect changes in attitudes and behaviours towards STWM of the target groups
Evaluation Purposes	**Effects** Review the engagement, behavioural, empowerment and delivery models used in the initiative. Evaluate the outputs of the project, focusing on the website and database. Assess the level and profile of participating organisations and individuals. Examine the types of projects securing funding. Evaluate the effects of participating in the initiative on awareness and attitudes towards science, on learning and on career decision-making. **Operational** Assess the extent to which the management and administrative systems of the initiative are 'fit for purpose' Evaluate the effectiveness of the partnerships working in the initiative **Accountability** Assess the cost-effectiveness of the initiative

Key Stakeholders	Large organisations with influence on science and technology education such as the DfES, the Science Museum, the British Association, The Royal Society. BME organisations including the African Caribbean Network for Science and Technology. Schools taking part in projects.
Evaluation Questions	**Effects** Are the engagement, empowerment and attitude and behaviour change models used appropriate for the purposes of the initiative? What has been the level of participation in the initiative? What kinds of schools and organisations have participated in the initiative? What have been the benefits for participating schools and organisations? What problems have been encountered and how have they been addressed? What kind of projects have been funded and what have they achieved? **Operational** How efficient and effective have the workshops and website been in supporting the scheme? How efficient and effective are the management and administration systems and procedures? **Accountability** Is the scheme value for money, set against its objectives?
Evaluation Criteria	**Effects** Conceptual coherence and appropriateness of empowerment, change and implementation models. Participation rates. Accessibility of the initiative. Relevance, quality and effectiveness of projects funded. Technical effectiveness and usability of website. Awareness of STEM issues. Learning outcomes. Self-reported career strategies. Science literacy scores. **Operational** Fitness for purpose of management and administrative systems. Efficiency and effectiveness of systems. **Accountability** Cost-effectiveness.
Methods	**Effects** Content and discourse analysis of project deliverables. Statistical analysis of participation data. Audit of funded projects. Case studies of selected projects. Stakeholder interviews. Schools assessment (treatment-comparison design). **Operational** Content analysis of minutes of meetings. Interviews. **Accountability** Cost-effectiveness analysis.

7.8. EXAMPLE 6: Prometheus

Description:

PROMETHEUS is a major government-sponsored programme aimed at supporting the implementation of policy initiatives such as the 'Investment Framework' – Science and Innovation Investment Framework: Next Steps 2004-2014, and the Science, Technology, Engineering and Mathematics (STEM) Report. Reflecting an investment of just over £15 million over a five-year period, the main objective of the programme is to support innovation in Science and Society research and development. The programme will fund around 30 individual projects in six thematic areas: public engagement; emerging technologies; indicators and benchmarking; professional development; access and accessibility; and communities and government. The programme incorporates a number of 'horizontal' actions to support the projects, including concertation; dissemination and exploitation and evaluation. It will be delivered via a range of innovative instruments, including co-funding arrangements involving the participation of the private sector.

Main objective:

To support innovation in Science and Society through targeted research and development.

Type:

Participation; support; awareness-raising; educational.

Duration:

Long (5 years).

Delivery:

Variable.

Scale:

Large.

Target groups:

Citizens.

Excluded groups

The private sector.

Voluntary organisations.

Communities.

Scientific societies.

Academics.

Schools.

Expected outcomes:

Improve and deepen understandings of how scientific discourses are created and sustained.

Improve and deepen understandings of the barriers to learning and career decision-making.

Develop innovative ways of raising awareness of scientific issues.

Develop innovative ways of engaging more people in STEM subjects and STEM careers.

Expected impacts:

Increase scientific capacity.

Create sustainable partnerships between industry, government, the voluntary sector.

Promote increase in numbers studying STEM subjects in education.

Improve national skills base in STEM sectors.

Innovation and complexity:

Highly innovative and experimental. Complex partnership and organisational arrangements.

Position in Delivery chain:

Covers spectrum of delivery chain.

Evaluation Approach:

The substantial level of investment, large scale and sustained duration of the programme, together with its broad scope, purposes and objectives and its inherent innovative, experimental nature, demands a multi-dimensional and multi-purpose evaluation design. The programme would benefit from building in evaluation at the planning stage, through ex-ante activities that would help the programme funders, managers and stakeholders to explore and optimise different design and implementation strategies prior to the launch of the programme. This developmental evaluation stance should be continued throughout the programme life cycle, so that a key purpose of the evaluation should be to work with stakeholders to feed ongoing evaluation results into the evolution of the programme, through regular feedback and reflective activities. Part of this process could include the elaboration of a 'theory of change' for the programme, together with a 'logic model' that could establish initial baselines against which subsequent programme and project outputs could be assessed. The developmental orientation should in turn be supported by evaluation activities that contribute to the operational effectiveness and efficiency of the programme on a continuing basis, through building linkages with the programme management. The programme allows for the design and implementation of evaluation as a 'horizontal' activity, in conjunction with the existing horizontal activities proposed on concertation and dissemination. This horizontal activity could include, inter alia, the production of guidelines and handbooks to support self-evaluation by the projects; an evaluation help desk and ad hoc evaluation consultancy. The 'summative' (effects) aspect of the evaluation needs to encompass the programme outputs, its outcomes and its impacts. These need to be assessed both from the perspective of the programme as a whole and from the perspective of the constituent projects and support actions. Evaluation resources should be devoted to carrying out longitudinal studies of impacts, focusing on attitudinal and behavioural changes associated with the programme, using appropriate projects as case studies.

Table 7-7 Prometheus

Evaluation Object	The programme, its architecture and its constituent projects.
Evaluation Purposes	***Planning*** To support the design and implementation plan for the programme. To identify and articulate its 'theory of change' and 'logical model'. To set up horizontal evaluation support. ***Developmental*** To provide evaluation support for programme managers and stakeholders. To support the evolution of the programme. ***Effects*** To assess the extent to which and in what ways the programme has achieved its objectives. To assess the contribution the programme has made to supporting policy. To assess the contribution the programme has made to enhancing the knowledge base and state of the art in the field. To assess the contribution the programme has made to supporting new practices. To evaluate the benefits for stakeholders of participating in the programme. To evaluate the outputs of the programme, and the projects involved. To evaluate the outcomes associated with the constituent projects, and the programme as a whole, in terms of their objectives.

	Operational Assess the extent to which the management and administrative systems of the programme are 'fit for purpose'. Evaluate the effectiveness of the partnerships working in the initiative. Provide ongoing evaluation results to improve the efficiency and effectiveness of the programme architecture. ***Accountability*** Assess the cost-effectiveness of the programme.
Key Stakeholders	Government Departments (DTI;DfES). Scientific societies. Academics and experts in the field. Commercial partners. 'Third sector' partners. Large organisations with influence on science and technology .
Evaluation Questions	***Planning*** What is the overall 'mission' and core vision of the programme? What are its 'theories of change'? What logical model best describes its objectives, resources and expected outcomes? What evaluation resources and support are required for the programme and its constituent projects? ***Developmental*** What problems are being encountered as the programme evolves and how can these be addressed? Are there ways in which stakeholders can be more actively engaged and their efforts more effectively valorised? ***Effects*** What kind of projects have been funded, what are their outputs and what have they achieved? What contribution has the programme made to supporting policy in the field? What contribution has the programme made to expanding the knowledge base in the field? What contribution has the programme made to improving practice? What transferable learning has the programme achieved? What are the main likely outcomes of the projects and the programmes in terms of key policy agendas, such as increasing the numbers of people taking STEM subjects and increasing the national skills base? Has the programme had any unforeseen or unintended effects (for example, substitution, displacement effects)? ***Operational*** How efficient, effective and equitable is the project selection and funding process? How efficient and effective are the management and administration systems and procedures? What problems have been encountered and how have they been addressed? How efficient and effective are the horizontal support actions? ***Accountability*** Does the programme provide good value for money, set against its objectives?
Evaluation Criteria	***Effects*** Comprehensiveness and equity of project representation. Participation rates and profiles of target groups covered. Coverage and relevance of policy areas covered. Degree of innovation.

	Relevance, quality and effectiveness of projects funded. Benefits to target groups. Transferability of programme outputs. Programme impacts. ***Operational*** Fitness for purpose of management and administrative systems. Efficiency and effectiveness of systems. Efficiency and effectiveness of partnership arrangements. Efficiency and effectiveness of funding instruments. ***Accountability*** Cost-effectiveness.
Methods	***Planning*** Stakeholder mapping. Concept mapping. Logic model. Critical reference groups. Expert Panel. Literature Review. Theory of change assessment. ***Developmental*** Action learning sets. Reflective Review Workshops. Self-evaluation guidelines and handbooks. Evaluation help desk. ***Effects*** Content and discourse analysis of programme and project deliverables. Statistical analysis of participation data. Audit of funded projects. Case studies of selected projects. Stakeholder interviews. Longitudinal studies. ***Operational*** Content analysis of management and administrative outputs. Stakeholder interviews. Focus groups. ***Accountability*** Cost-effectiveness analysis.

8. ILLUSTRATION: EVALUATION APPROACH FOR 'SCIENCEWISE'

8.1. What this section is about

Following on from the practical examples set out in Section 7, this section provides a more in-depth case study of an evaluation of a real world Science and Society programme. It provides a detailed, step-by-step illustration of how to design an evaluation in practice.

In this final section, we sketch how the Evaluation Framework could be applied to design an evaluation of a real-world OST activity – the 'Sciencewise' initiative.

8.2. Main Objectives of the Evaluation

The overall objective of this research is to carry out an evaluation of the OST 'Sciencewise' programme. The Sciencewise programme sits within OST's broad 'mission', which is to invest in the promotion of public engagement in science and in developing and promoting Science and Society initiatives. Against this background, the evaluation is intended to help OST gain an understanding of:

- The extent to which Sciencewise informs and challenges current policy and practice in science policy in general and in Science and Society type initiatives.
- The contribution it makes to promoting innovation in these policy and practice areas.
- What impact the programme has (on 'beneficiaries' – i.e. participants funded; audiences addressed) – on policy-makers and on deepening and expanding the knowledge base in science (e.g. in the practitioner community).
- Whether the way the programme is managed is 'fit for purpose' (e.g. selection of particular funding; monitoring of the projects).

And

- Help identify examples of good practice and 'success' to provide inputs for OST profile-raising activities.
- Help make short-term evolutionary improvements to the programme.
- Inform strategic decisions about the longer-term future of the programme.
- Apply what has been learned to the development of OST's other programmes and the development of new programmes.

The main ***questions*** the evaluation addresses are:

- How fit for purpose is the programme model? This reflects a number of subsidiary questions, including:
    - o What is distinctive about the programme within the Science and Society landscape?

- Is the programme sufficiently focused to have an impact on key strategic areas of science and public engagement in science?
- Has the programme been able to source innovative projects?
- Is there a link between the programme's flexibility and the quality of proposals submitted and funded?

- What impact has the programme had to date? This also reflects a number of subsidiary questions, including:
  - To what extent have the projects funded benefited their direct beneficiaries?
  - What is the quality of evidence and transferable learning produced by projects? Has the approach to project evaluation been robust?
  - What potential lasting benefits have been generated from the partnerships the programme has developed? What can be said about the sustainability of projects funded?
  - Has the learning from the projects and the programme been used to inform Science and Society policy?

8.3. Evaluation activities and workplan

8.3.1. Overview

The evaluation design incorporates seven inter-related sets of activities (work packages):

- WP1: Scoping
- WP2: Process Review
- WP3: Secondary data analysis
- WP4: Case studies
- WP5: Comparisons and context
- WP6: Analysis & Report
- WP7: Project Management

8.3.2. Workplan

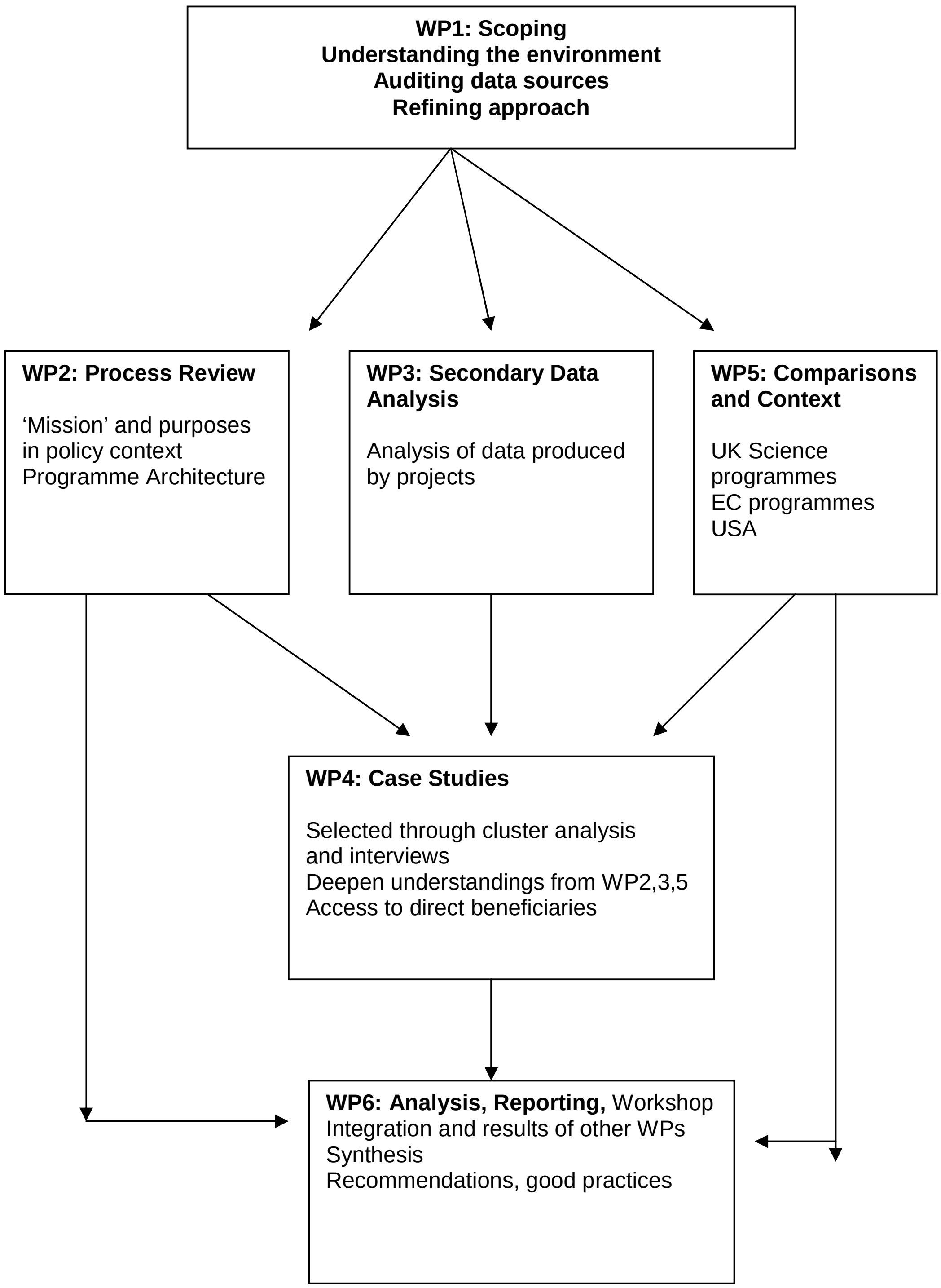
Figure 8:1 shows how these elements of the workplan are integrated within the overall evaluation. As the figure illustrates the starting point for the evaluation is a Scoping Exercise (Workpackage 1). This will enable the evaluation team to: gain an understanding of the policy environment in which OST, and in turn Sciencewise, operates; carry out an 'audit' of relevant data sources and contacts; recruit an 'Expert Panel' to provide inputs to later evaluation tasks; 'fine tune' the evaluation approach and methodology; and develop a data collection 'toolkit. The methods used will involve: review of relevant documentation (e.g. OST consultation documents; Sciencewise planning documents) and interviews with key actors. The table below shows the activities involved.

What	How
Prepare detailed methodology	Methodology Report
Preliminary interviews with programme managers	Structured interviews
Review of all available documentation (files) on the projects	Content Analysis (Consultation documents; Review; Committee Minutes)
Other background interviews	Expert Panel set up

Following the scoping exercise, three sets of evaluation activities will be undertaken, broadly in parallel. These are:

- Process Review (work package 2)
- Secondary Data Analysis (work package 3)
- Comparisons and Context appraisal (work package 5)

Figure 8:1 Inter-relationship of workplan elements



The ***Process Review*** has three main purposes:

- To review the mission, vision and purposes of Sciencewise within the context of OST's key purposes and those of its 'sister' programmes.
- To assess the coherence, relevance and contribution of the programme within the broader context of current and future policy and practice agendas in learning.
- To review the 'architecture' of the programme (including its selection, funding, monitoring and evaluation procedures).

The methods used in the process review will include: Expert Panel assessment; interviews / focus groups with OST staff and other key stakeholders – e.g. government agents – (face-to-face / telephone / email); 'Cultural logic' analysis (discourse / content analysis) of key documents.

The table below shows the activities involved.

What	**How**
Group / individual interviews with project managers	On-line survey Case Study interviews
Interviews with OST staff and others involved in the project selection / funding process	Structured interviews
Assessment of Programme Architecture	Content Analysis Structured interviews
Policy analysis / cultural logic analysis	Content Analysis Structured interviews

Our analysis of ***Secondary Data*** is intended to:

- Extend our 'background picture' of the programme by assessing the main features of the spectrum of projects funded by the programme.
- Explore the similarities, differences and inter-relationships between projects.
- Identify a sample of projects as candidates for the more detailed analysis carried out in the case studies.

The data capture methods used will include interviews with OST staff and project staff (face-to-face / telephone / email); content analysis of available documentation and website content. The analysis will focus on a 'cluster analysis' of funded projects using a set of key variables to develop a 'typology' of projects (for example, on the basis of geographical distribution; distribution of funding; pedagogic model; type of outputs; target beneficiaries; innovation type). The table below shows the activities involved.

What	**How**
Desk research on characteristics of projects funded	Project audit – content analysis of profiles
Analysis of funded projects	On-line survey of project managers Results of content analysis Analysis of OST data
Meta-analysis of evaluation	Content analysis of case study reports

reports produced by projects	
Citation analysis	Search ERIC BEI bibliographic databases Google Scholar search Search sample of on-line journals

Alongside the above activities, we will undertake a ***Comparisons and Context appraisal***. This is intended to situate the programme in relation to comparable 'benchmarks'. These might include: other OST programmes; UK Science and Society Programmes (e.g. ESRC; 'Big Lottery; New Deal for Communities); European Commission Programmes (e.g. COMENIUS); Information Society Technology; eLearning; Minerva. The methods used will include: Expert Panel input; synthesis of existing evaluation reports and relevant documentation; and comparison of available statistical data.

The table below shows the activities involved.

What	**How**
Literature and documentary review	Database search Interviews with Expert Panel
Interviews with external key informants	Expert Panel on-line Forum

The results of these three activities will feed into work package 4 – ***Case Studies*** – which is the main platform through which the outputs, outcomes and potential impacts of the programme will be assessed. The majority of these will be in the form of studies of particular projects – 'cases' – that are chosen as 'exemplars' or typical examples of the clusters identified in work package 3. At least one of the cases will take the form of a 'thematic' case study – for example, illustrating the life cycle of the programme and how it has evolved. Each case study will entail the use of a 'template' (common data collection and analysis 'toolkit') to promote standardisation of data and enable cross-case comparison. The methods used in the toolkit will include: self-administered questionnaires (SAQs); observation; secondary data analysis; focus groups; and content analysis of documentation.

In response to time and resource constraints, the strategy adopted in the case studies will be to optimise data collection efficiency and effectiveness by minimising the need to carry out time-consuming (and difficult to arrange) primary data collection by the evaluation team. Thus, when possible, primary case study data will be collected through 'alternative' means, including:

- 'On-site' data collection (using observation; quota sample SAQ) – e.g. in the Natural History Museum.
- 'Intermediary' data collection (using, for example, SAQs implemented via project staff).
- 'Automated' data collection, focusing on those projects with distinctive 'technology' delivery approaches (using, for example, on-line SAQ; weblog analysis).

The table below shows the activities involved.

What	How
Selection of indicative sample of projects	Cases selected on basis of audit data
Field work	Observation Interviews with project co-ordinators SAQs with beneficiaries Focus groups Content analysis of project reports, media and outputs
Results synthesis	Case study template Cultural logic analysis
Review of Results	Expert Panel Forum. The Forum will be given a summary of the results

The results of the case studies, together with those of the preceding work packages, will be analysed, integrated and synthesised by the project team, working in collaboration with OST staff. A dedicated project management work package will monitor the progress of the evaluation throughout its life cycle.

ANNEX I: INNOVATIVE METHODOLOGY

In this Annex we present a number of ‘innovative’ methodologies: ex-ante evaluation; and longitudinal or prospective research. We use the term ‘innovative’ here, not in the sense that there is anything new about the methods discussed, but in the sense that they have been relatively under-utilised in the Science and Society field. The two methods described below have been chosen in collaboration with the commissioners of this project, as it is felt that they offer particular promise for addressing questions of interest to funders of Science and Society initiatives.

Ex-Ante Evaluation

Ex-ante evaluation is often used in the kinds of large-scale programmes referred to in EXAMPLE 6: Prometheus above – typically those involving significant public investment (for example, European Union structural funds). Ex-ante evaluation is commonly understood as a way of designing programmes that incorporate some vision of the kinds of future outcomes and impacts they are intended to create.[18] However, ex-ante evaluation often involves the implementation of both a specific ‘theory of change’ model as well as instruments that are designed to assess the extent to which the model’s intended outcomes and impacts are realised (Weiss, 1995).[19] ‘Theory of change’ approaches in evaluation, when set within the context of ex-ante evaluation, typically involve key stakeholders in the design of the ‘theory’ behind interventions. Sullivan and Stewart, for example, review this type of ex-ante approach in relation to three major UK public initiatives: Health Action Zones, New Deal for Communities and Local Strategic Partnerships. [20]

Ex-ante evaluation takes place at the beginning of the cycle before a programme has been adopted. This form of evaluation helps to ensure that the final programme is as relevant and coherent as possible. Its conclusions are intended to be integrated into the programme when decisions are taken. It also provides the required foundations for monitoring and for future evaluations by ensuring that there are explicit and, where possible, quantified objectives. It helps to specify selection criteria for the selection of projects within a programme and to ensure that policy priorities are respected. Finally, it helps to ensure the transparency of decisions by allowing for a clear explanation of choices made and their expected effects. Evaluation methods commonly used in ex-ante evaluation include: stakeholder mapping and analysis; concept mapping; logic models; Delphi methods; and road mapping.

[18] See, for example ‘Ex ante evaluation of options for development of a competitive, dynamic and sustainable knowledge society 2006-2013. (2006) European Commission, DG Information Society and Media, Brussels

[19] Weiss, C 1995. "Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families." In *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*, ed. James P. Connell et al. Washington, DC: Aspen Institute.

[20] Sullivan H and M Stewart (2006), Who Owns the Theory of Change? Evaluation, Vol. 12, No. 2, 179-199

Stakeholder mapping and analysis

Stakeholder consultation is a participative technique for involving stakeholders in policy and programme evaluation. The technique may be used at a specific stage of the evaluation, most commonly in identifying evaluation priorities and questions at the outset of the evaluation. Data collection methods include surveys of stakeholders by interview or questionnaire. Respondents provide information about their perceptions of the programme being evaluated, their strategic interests and priorities, the kinds of questions they want answered and how evaluation findings can best inform their decision-making. They may also be asked to prioritise their responses or to indicate the relative importance of different items.

Concept mapping

Concept mapping is used to define the effects that are to be evaluated and, in cases where there are multiple objectives and where these have not yet been firmly established, the associated indicators. It may also be very useful if the set of objectives lack precision. The tool is suited to evaluation in a partnership context because it is based on the aggregation of individual points of view for the purpose of reaching consensus between the partners.

Logic models

Logic models are part of a wider family of evaluation approaches that seek to describe programme theory. The logical framework technique is an exercise in structuring the component elements of a project (or single programme) and analysing the internal and external coherence of the project. The product of this technique, the logical framework, is a formal matrix presentation of the internal functioning of the project, of the means for verifying the achievement of the goal, and of the internal and external factors conditioning its success.

Delphi method

The Delphi method is based on a structured process for collecting and synthesising knowledge from a group of experts by means of a series of questionnaire surveys accompanied by controlled opinion feedback (Adler and Ziglio, 1996). Delphi is primarily used to facilitate the formation of a group judgement (Helmer, 1977). It developed in response to problems associated with conventional group opinion assessment techniques, such as focus groups, which can create problems of response bias due to the dominance of powerful opinion-leaders (Wissema, 1982). In ex-ante evaluation it may be used in forward planning to establish hypotheses about what programmes are intended to achieve and how they can be implemented. This technique may be used when significant expertise exists on the subject, but where new experiments are being developed to expand the knowledge base.

Road mapping

Road mapping is not an additional or alternative research method to Delphi – indeed, Delphi surveys are often used to provide inputs to developing roadmaps. This reinforces the widely accepted view that road mapping moves forward from an analytical and interpretative perspective to provide both a prescriptive and operational framework. Motorola first introduced the concept of a 'roadmap' in the 1970s as a kind of strategic planning tool, and these origins reflect two of the key characteristics of the approach: its close association with technology forecasting and its almost exclusive use, until recently, as a framework and vehicle for designing and developing change programmes within large organisations[21]. In simple terms, a roadmap can be understood as a time-based plan that defines where an organisation is, where it wants to go and how it can get there. This includes identifying precise objectives and paths for meeting certain performance objectives and helping to focus resources on the critical elements that are needed to meet these objectives. Roadmaps are both forecasts of what is possible or likely to happen, as well as plans that articulate a course of anticipatory action. In all roadmaps, the future is expressed in terms of a number of scenarios – which may be alternating or complementing. In the case of ex-ante evaluation, these scenarios are both retrospective, in that they embody evidence based on past experiences, and prospective, in the sense of presenting 'possible futures' or trajectories of these past realities. In programme design, these scenarios may reflect 'thematic' elements of the programme, as in EXAMPLE 6: Prometheus above.

Prospective evaluation

Ex-ante research, and the tools described above, are part of a genre of 'prospective' evaluation and research approaches and methods. In general, prospective methods are 'interpretive', in the context of ex-ante evaluation, since they typically involve eliciting and mapping opinions. However, one particular strand of prospective evaluation entails using classic 'experimental' methodologies. In the context of the experimentalist research tradition, prospective research is generally accepted to be a procedure involving longitudinal studies aimed at analysing the effects over time of a phenomenon or phenomena on the 'units of analysis' or cases being studied (i.e. generally speaking, people). One expert defines it in the following terms:

> *"A research strategy in which people are followed forward in time to examine the relationship between one set of variables and later occurrences. For example, prospective research can enable researchers to identify risk factors for diseases that develop at a later point in time".* [22]

As implied by the above definition, most experimental prospective research has been carried out within the health and medical domains, frequently involving randomised controlled trials (RCTs) of drugs and other therapies, across a wide range of conditions, from oncology to coronary illness. [23] [24] Not all applications of prospective research involve clinical trials. Other studies have adopted an exploratory and analytical perspective rather than a 'control-comparison' position. Examples include

---

[21] Roadmap to Communicating Knowledge Essential for the industrial environment, IST Programme

[22] Taylor , S (2003) 'Health Psychology', Los Angeles, McGraw Hill
[23] For example: Illing HM, Morris DO, Lee RT. (1998) A prospective evaluation of Bass, Bionator and Twin Block appliances. Part I—the hard tissues. *Eur J Orthod* 1998; **20**: 501–516.[

[24] Reynolds JV, McLaughlin R, Moore J, Rowley S, Ravi N, Byrne PJ (2006). Prospective evaluation of quality of life in patients with localized oesophageal cancer treated by multimodality therapy or surgery alone. Br J Surg. 2006

research into how and why adolescents take up smoking,[25] studies of sexual behaviours [26] and research into the causes and effects of child abuse. [27]

The common element that binds together these different topics is a desire to understand a phenomenon not by taking it apart and trying to work out what its constituent parts do, but by analysing it over time. In other words, experimental prospective research has an ***action-oriented*** focus. It involves evidence-based research (usually encompassing RCTs) that aims to establish the causal relationships and the effects of the phenomena being studied rather than the ontological and epistemological properties of the phenomenon itself. This in turn requires the development of an initial ***process model*** and its subsequent refinement on the basis of evidence collected from ongoing studies over time. Take the example of adolescent contraceptive use. This involved a longitudinal study of a cohort of 375 low-income African–American adolescent girls over a period of six months. At the start of the study (time 0) each participant was given a urine test to test for pregnancy and was asked to complete a self-administered questionnaire on sexual behaviours, attitudes to contraception and attitudes to pregnancy. Six months later, a follow-up pregnancy test was administered to the same cohort, together with a similar questionnaire on sexual and contraception behaviours and on attitudes to pregnancy. A regression model was developed at the outset of the study to calculate odds ratios of inconsistent contraception (i.e. the model identified which members of the cohort were less likely to adopt contraception strategies on the basis of 'baseline' predictors like income). This model was then compared with the actual behaviours of the cohort on the basis of the evidence collected.

The main result of the study was that adolescents who were inconsistent contraceptive users at follow-up were more likely to have reported a desire for pregnancy, previous inconsistent contraceptive use, less frequent communication with their partners about prevention issues and an increased number of lifetime sexual partners at the baseline assessment. Of equal importance was the finding that a previous pregnancy or sexually transmitted infection did not influence future contraceptive behaviours.

This example shows that 'experimental' prospective studies:

- Incorporate an initial process (action) model that is used to i) specify the relationships, interactions and variables that will be studied ii) predict the behaviours and outcomes of the study population iii) deepen understandings of the phenomenon being studied by refining the process model on the basis of the 'goodness of fit' between predicted and actual outcomes.
- Deploy a 'pre-test-post-test' research design.

[25] Pedersen W, Lavik NJ. (1991), Role modelling and cigarette smoking: vulnerable working class girls? A longitudinal study. Scand J Soc Med. 1991 Jun;19(2):110-5.

[26] Predictors of inconsistent contraceptive use among adolescent girls: findings from a prospective study. J Adolesc Health. 2006 Jul;39(1):43-9.

[27] Beutler, L.E., and Hill, C.E. (1992) Process and outcome research in the treatment of adult victims of childhood sexual abuse: Methodological issues. *Journal of Consulting and Clinical Psychology* (1992) 60:204–12

- Involve a cohort drawn from the study population who participate in a real-time, ongoing data collection process.
- Can be used to develop prescriptive strategies that are rooted in an evidence base derived from examination of real behaviours. In this case, the findings have significant implications for the development of effective sexually transmitted infection (STI) and pregnancy prevention programmes for adolescents and can help in guiding clinicians toward relevant treatment practices.

In the case of EXAMPLE 6: Prometheus, described above, the use of concept mapping, theory of change analysis and logic models would provide the basis for developing a process model for the Prometheus programme that could subsequently be used to develop longitudinal studies that could test the model.

However, there are a number of problems associated with the 'experimental' approach and the use of methodologies including longitudinal studies, randomised control trials and treatment-control evaluations. These are discussed below.

Experimental approaches

Description of the method

Experimental approaches are often described as the 'gold standard' of evaluation (Cambell and Stanley, 1973). Essentially, they attempt to replicate the kinds of conditions, in a 'social' context, under which the behaviours of so-called 'natural science phenomena' – such as atomic particles – are observed and understood in the laboratory. The most commonly used experimental method is the randomised control trial (RCT). Most people encounter RCTs in everyday life in the form of 'clinical trials' – testing the effectiveness of a new pharmaceutical product in a sample of an appropriate target population, for example. The philosophical basis for this approach has its origins in positivistic and scientific-realist traditions that began to develop during the Enlightenment, and which were consolidated during the eighteenth and nineteenth centuries, as 'science' and 'technology' began to exert an increasingly powerful control on the production of knowledge. The central assertion of this philosophical tradition, that 'truth' can only be established by testing it empirically in the real world, underpins the methodological 'rules' or 'protocols' under which RCTs operate.

A commonly-used definition of RCTs (Schwarz et al, 1980) illustrates the characteristics of this experimental world-view:

> *"(An RCT is) a prospective experimental study in which the effects of one or more interventions are assessed by dividing a research population on the basis of random allocation into one or more experimental and one or more control groups."*

Random allocation means a process of allocating each unit of analysis to either the experimental or control groups so that each unit has an equal chance of receiving the 'treatment'. 'Units of analysis' are typically 'people', although the research population can more accurately be defined as the 'outcomes' of the treatment (for example, behaviours or attitudes measured). A 'treatment' (the intervention) is administered to the experimental group, but not to the control group. The effects of the treatment are comparatively assessed between the experimental and control groups in order to judge the extent to which the treatment has had a significant effect on the experimental group in terms of some common measurement standard.

Purposes of the method

The main purpose of experimental techniques such as RCTs is to gather enough reliable evidence about a particular phenomenon in a particular situation to enable the behaviour of that phenomenon to be predicted in similar situations. They, therefore, attempt to establish causal relationships and laws. From a technical standpoint, this means that the purpose of carrying out a randomised control trial is to verify that the following four conditions hold true: validity; reliability; predictability and temporal priority. Temporal priority is the raison d'être for experimental trials. It is shorthand for the assumption that a suspected cause precedes an event (for example, in clinical trials that the application of a particular drug will 'cause' the relief of particular symptoms). In order to demonstrate temporal priority, RCTs must be able to *control* for the influence of intervening variables in order to isolate the cause of an effect. Failure to isolate the effects of variables other than the 'treatment' variable will affect both the *internal validity* of the trial (the measurability of the relationship between the treatment and its outcomes; i.e. that the trial is measuring the effects of the 'treatment' and not some other 'intervening' factor) and its *external validity* (the generalisability of the outcomes of the trial to other settings). The validity of the trial will in turn dictate its *reliability* (whether the trial produces the same results when repeated) and *predictability* (the extent to which the results can predict outcomes that may occur in similar situations).

Random assignment of units of analysis to experimental and control groups allows the opportunity to demonstrate temporal priority whilst dispersing the effects of intervening variables. It utilises the properties of probability distributions in order to avoid biasing the results of trials, for example, as a result of concentrating subjects with common characteristics in one group. At the same time, it allows for reliability and predictability by utilising the properties of population distributions in terms of sampling theory: if the randomly-assigned population is normally distributed, then the outcomes of the treatment can be extrapolated to predict the effects of the treatment on other populations with similar characteristics.

Circumstances in which it is applied

Debates about whether or not to use an experimental approach, such as a Randomised Control Design for the evaluation of a social programme can generate strong feelings on both sides. On one side, the proponents of the design argue that it is the only way in which the causal relationships which are assumed to be the basis of a particular intervention can be scientifically proved. Without subjecting a new intervention to this test, it is argued, we are using methods which, at best, may be wasteful of limited resources and, at worst, harmful to those on which it is being used.

On the other side, those who favour more qualitative or exploratory research methods, sometimes also the participants in the programme itself, fear that the methodology of the RCT will fail to capture the complexity and uniqueness of the activities being researched. Given this limitation, it is argued, there is a danger that the results, even if statistically significant, will fail to contribute in a real way to the understanding of the effectiveness of the intervention under different circumstances.

Meanwhile, it is sometimes argued that discussions about the appropriateness of an RCT often do not go far enough; that what needs to be considered is not just whether an RCT is the appropriate methodology, but whether the research paradigm, within which the RCT methodology is located, is the right one for the situation under consideration. As Guba and Lincoln put it:

> *"We have argued often that evaluators who operate in a constructivist, responsive and now fourth generation mode will use primarily although not exclusively qualitative methods. But there will be times when quantitative methods – tests or other measurement instruments or numeric displays – will be, and should be used. The single limitation that a constructivist, responsive, fourth generation evaluator would put on the use of quantitative methods is that no causally inferential statistics would be employed since the causal linkage implied by such statistics are contrary to the position on causality that phenomenologically oriented and constructivist inquiry takes." (Guba and Lincoln 1989)*

As Kuhn originally asserted (Kuhn, 1962), a scientist/researcher operates within a particular paradigm, not only because of its intrinsic content, but also because of the way in which they have been trained and socialised in their particular field of research. A number of forces in recent years have encouraged social researchers to question the limitations of research paradigms which lie within the experimental tradition. These include, the growing demand for research to be empowering for those disadvantaged groups that are frequently its 'subjects', and developments in feminist research and research in the developing world.[28] The experimental tradition, it is argued from a number of sides, is a product of a particular cultural and historical environment, and can lead to a systematic and unrecognised bias in those undertaking research within that tradition. Against this background, there has been a strong movement within the field of evaluation towards research approaches which accept the multiple nature of social realities and which can incorporate the dimension of power and inequality between different participants (Parlett and Dearden, 1977; Guba and Lincoln, 1989). This movement encourages evaluation activities in which consultation with different 'stakeholders' in the activity under evaluation takes a central place, where understanding and consensus building is a more important outcome from the evaluation than the identification of one, validated 'explanation', and in which considerable emphasis is placed on the empowerment of stakeholders whose views, within a more conventional research design, may have been overlooked.

More prosaically, as discussed below, the limitations of RCTs, particularly in the field of social research, are often expressed in terms of practical problems in applying the methodology in a social setting. Other problems cited include ethical issues and the cost of implementing trials in resource terms.

---

[28] See, for example, the works of , Friere (1970), Gordon (1975), Ardener (1975) and Oliver 1990.

In the light of these complex and difficult issues, it is suggested that a decision on when to use an experimental technique needs to be shaped by the following questions:

- What are the purposes of the evaluation?
- What research questions are being asked?
- What is the environmental context of the evaluation?
- What is the operational context of the evaluation?
- What are the purposes of the evaluation?

In relation to ***purposes,*** experimental techniques tend to be inappropriate in exploratory, descriptive or participatory and action research evaluation environments. They are more suitable in situations where the main evaluation purpose is:

a) Testing or validation of a model, focusing on assessing how the intervention works and whether the choices embedded in the model are acceptable to different stakeholders.

b) Experimental, predictive and/or hypothesis-testing purposes – for example, comparison of the effectiveness and outcomes of the intervention in different field settings.

In relation to ***questions,*** experimental techniques are unlikely to work in situations where the research questions are:

- Exploratory, for example, identifying the main issues in a domain where little is already known.
- Descriptive, for example, elaboration identified main issues as input to hypotheses.
- Critical, for example, questions on the political issues around an intervention and whether there is consensus about it.

Experimental techniques are more likely to be employed in situations where the research questions are:

- Explanatory, for example, to establish causal linkages between variables.
- Predictive, for example, to test out what is likely to happen if an intervention is applied elsewhere.

In terms of the setting or environmental context of the evaluation, experimental techniques tend to be relatively poor ways of capturing belief systems, symbolic meanings and the ways in which interventions are re-created and adapted in interaction with social life. Similarly, they are not very useful, for example, when looking at how the intervention has been shaped by historical trends and social forces, and how the intervention is likely to evolve and be shaped by such forces in the future. This implies that experimental techniques should not be used in situations which involve the interpretation of 'multiple constructions of reality' – or where there is likely to be a range of radically different world views amongst stakeholders.

In making a judgement about the appropriateness of experimental techniques and running randomised controlled trials, consideration needs to be paid to the logistics or ***operational context*** of the evaluation. The key elements that need to be considered are:

- The methodological preferences and ‘world view’ of stakeholders.
- The resource requirements involved.
- Available skills and expertise to implement the trial.

Review who has a stake in the intervention assessment, particularly the main clients, funders of the research and consumers of the research outputs. This review needs to ask what type of research approach would these stakeholders find credible, and is the methodological orientation of key stakeholders compatible with the experimental paradigm of RCTs? Review the resources available for the assessment, bearing in mind the fact that, as discussed above, RCTs can be time-consuming, labour intensive and difficult to analyse. This is particularly relevant with trials involving complex permutations of sub-groups within treatment and control groups, and for longitudinal studies. Consider who will implement the trials and establish whether there is: the organisational capacity to implement them (especially in situations involving several sites); the relevant skills available (in terms of research design, field work and data analysis); and appropriate data management capability (ensuring quality controls in data collection and analysis; storage and retention of collected data and results), since experimental techniques imply a high degree of data integrity.

The main steps involved

Step 1: Assessment of the appropriateness of experimental approach to the evaluation scenario.

As discussed above, there are many situations in which the experimental paradigm itself may not be appropriate for the intervention to be evaluated. In order to establish this, the following actions need to be undertaken:

- A stakeholder audit (to identify the ‘world views’ of the different stakeholders; their receptivity to different types of data analysis; the degree of ‘politicisation’ of the intervention).
- Obtain answers to the four key questions specified above, i.e.
    - What are the purposes of the evaluation?
    - What research questions are being asked?
    - What is the environmental context of the evaluation?
    - What is the operational context of the evaluation?

Step 2: Assessment of the technical and methodological feasibility of the experimental design.

Although the experimental approach may be both appropriate and desired by the stakeholders involved, in many situations it is almost impossible to carry out in practice. In essence, the more complex the RCT design, the greater the investment in time, resource and analysis required. This is particularly acute in relation to involving sufficient numbers of units of analysis in the trials in order to conform to statistical rules governing normally distributed populations. It has been argued that experiments involving social interventions “simply cannot bear the expense of a sample size large enough to measure all of the possible combinations of treatment mix and client characteristics that characterise the operational environment of a program at a given point in time and as it changes over time” (Stromsdorfer, 1987).

The factors that typically conspire to undermine the integrity of an experimental design are discussed below under ‘strengths and limitations of the approach’. It is, therefore, good practice to review the methodological conditions under which the approach will be implemented with reference to these factors before making a decision to proceed. Should there be a significant possibility that the experiment is likely to be compromised, then better results are likely to be obtained by using a ‘quasi-experimental’ method – discussed in more detail below.

Step 3: Define hypotheses to be tested; identify key indicators for comparison.

Since the main objective of the experimental approach is to assess whether and in what ways exposure to a particular intervention or programme has a measurable effect, it is crucial to define beforehand which particular aspect of ‘temporal priority’ is being tested and on which particular criteria. Say, for example, the objective of the experiment is to assess the effectiveness of a ‘Scienceaware’ promotion pack in increasing young people’s participation in studying STEM subjects, a number of hypotheses might be generated for testing purposes. These could range from a prediction that awareness of scientific issues would increase in the ‘treatment’ population to a prediction that the actual STEM study rate amongst the treatment population would be increased. Different hypotheses imply different measurement criteria. In the former case, for example, the measurement criteria selected might be based on rating scales assessing the degree of knowledge of the sample population of scientific issues. In the latter case, measurement may encompass comparison of actual rates of STEM subject selection in schools from where the treatment group was drawn, compared with prevalence rates in similar communities elsewhere.

Step 4: Random assignment of the research population into groups and implementation of the experiment.

As discussed above, the essential aim of the experiment is to recruit two groups of participating ‘respondents’: those who are or have been exposed to the programme or intervention (experimental group) and a strictly identical group (control group) that allows what would have happened in the absence of the programme to be observed. (Note: the literature on experimental techniques sometimes creates confusion by using the term ‘control-comparison groups’ to denote ‘experimental-control groups’. In the former case the ‘control’ group is actually the ‘experimental group’.)

In practice, the basic randomisation design – known as the ‘Solomon four-group model’ – involves four ‘cells’, as shown in the table below.

A. Experimental Group	**B. Control group**
A1. Pre-test	B1. Pre-test

A2. Post-test	B2. Post-test

Imagine the experiment involves an evaluation of a ‘hot-housing’ programme on the learning skills of a cohort of primary school children in an inner urban community. The procedure would be to i) identify and recruit a random sample of, say, one hundred children ii) administer a common questionnaire schedule to all one hundred (assessing literacy skills; learning styles and so on) iii) divide the sample (again on the basis of random assignment) into two groups – experimental (or ‘treatment’) and control iv) run the programme only for the treatment group v) apply the questionnaire schedule again with both the treatment and the control group.

Step 5: Comparison of results.
At the end of the experimental period, the data collected during the implementation phase is compared across the treatment and control groups, as set against the hypotheses being tested and the measurement criteria used (as discussed above in Step 3). This will typically involve multi-variate analysis, such as regression analysis and techniques that compare measurements between and within groups (for example, students’ t-tests).

Strengths and limitations of the approach

As discussed above, the main strength attributed to experimental techniques is that they represent the most rigorous and ‘scientific’ approach to assessing outcomes and impacts, and of inferring ‘causality’ between interventions and outcomes and impacts. However, they are extremely difficult to implement successfully, mainly because of the operation of three factors: history effects, selection effects and instrumentation effects.

Take as an example the ‘Scienceaware’ initiative referenced above. This aims to help ‘hard to reach’ groups – such as white working class, black and female, young people to engage more fully in STEM activities – by providing them with support and information. The main objectives of the intervention are to:

- Raise awareness about science.
- Provide support in decision-making, for example, subject choice.
- Increase the level of individuals taking STEM subjects.
- Support career decision-making.

The evaluation method chosen is a randomised control trial involving a sample of tenants selected for participation in the scheme – the ‘treatment’ group, compared with a sample of people with similar demographic and lifestyle characteristics – the control group. However, an assessment of the effects of the programme on the behaviours of the two groups is likely to be inhibited by:

- *History effects* – which involve 'interference' either from external influences or maturation effects occurring 'internally' as a result of the passage of time, such as individual ageing. In this example, these could include the effects of other, unrelated STEM initiatives on the ‘control group’ sample.

- *Selection effects* – statistical bias resulting from the fact that the treatment group and control group, despite being randomly assigned, are in reality not statistically equivalent. This is quite likely to occur in this evaluation for two reasons – firstly, because members of ‘hard to reach’ groups are difficult to recruit for such initiatives and their evaluation, limiting sample choice; and, secondly, because the small size of the sample substantially increases the possibility of (and the implications of) differences between the two groups.
- *Instrumentation effects* – these occur where treatment effects are measured using different instruments or under different conditions for one group to the other group. This can happen, for example, where participants in the treatment group are paid some form of incentive to take part whereas those in the comparison group are not.

Two further problems with experimental techniques are particularly pertinent here: attrition and lead times. It is almost inevitable that randomised control trials haemorrhage subjects over the study and this *attrition* tends to be more pronounced for members of excluded groups. The Program of Assertive Community Treatment (PACT) example is fairly typical (Stein and Test, 1980). Over a one-year period, a randomised trial was carried out involving two hundred inmates of a large urban city jail. Over half the total subjects randomly assigned to the trial at its commencement dropped out before the end, with the drop-out rate most pronounced in the control group, which lost 73% of its original participants. This attrition was precipitated by a number of factors: some subjects refused their consent at the start or later on during the study (around 10%); others could not be located on release from jail (19%); others went directly from jail to long-term institutions (10%). Such attrition is likely to affect the validity of results of the trial mainly because of resultant non-comparability of subjects across conditions and because the rules governing the statistical models normally adopted to analyse the data (such as in classic linear regression and analysis of variance) tend to be contravened (Graham and Donaldson, 1993).

The long *lead times* involved in facilitating and measuring attitudinal and behaviour change constitute a major methodological problem. It is widely recognised that community level initiatives in areas such as healthcare, regeneration and adult learning are long term enterprises. For example, in a Healthy Living Centres seminar held in April 1998, one of the participants referred to the fourteen years that it took for local residents to take ownership of their own development opportunities. An implicit goal of many public initiatives in these kinds of fields is to seek to foreshorten longer-term development processes. This poses problems for evaluators seeking to demonstrate achievements and outcomes within the lifespan of an evaluation. At the very least it requires a life cycle framework that acknowledges intermediate outcomes as well as longer-term impacts, although this inevitably heightens the problem of establishing the linkage between different elements in the presumed chain of causality.

Quasi-experimental techniques

There are a number of ways of addressing these problems within the broad ‘experimental paradigm’, such as quasi-experimental methods and statistical modelling. Quasi-experimental methods relax the probabilistic and population distribution conditions imposed by ‘true’ experimental research designs by shifting the emphasis from ‘cause-effect’ in temporal priority to ‘association’ between variables.

Commonly used methods include time series designs and non-equivalent groups designs. The former involves periodic measurement of an experimental and control group over a given period of time during which a treatment is applied. The latter method is essentially the same as a classic pre-test post-test experimental design, except that subjects are not randomly assigned to the different treatment conditions. Matching of pairs of subjects from the experimental and control groups, on the basis of characteristics like gender and age, enables a degree of control over intervening variables to be built into the trial.

One strategy used is to draw on available evidence from studies already carried out in a particular domain in order to statistically model the behaviours of intervention target groups. Another is to carry out a survey of the comparison group prior to the onset of the trial proper. For example, suppose an evaluation of an employment-generation program for long-term youth unemployed is commissioned. For political reasons, it is not possible to implement a total randomised trial of the intervention. The evaluators approach a government agency, which has implemented a similar initiative in the past, and ask it to provide them with existing socio-demographic data on participants who had enrolled in the past initiative, together with a list of candidates for the scheme who were eligible, but who in the event did not participate. This list of candidates is then surveyed on a range of measures eliciting standard sociodemographic information, together with other information on attitudes towards employment. On the basis of the data on past participation, together with data derived from the survey, a statistical model identifying the main predictors of participation is developed, using probit analysis (Finney, 1971). This model enables clusters or typologies of participation characteristics to be identified, and can provide a means of selecting a comparison group which, in theory, will provide a 'true' representation of the target population rather than one which is biased towards a particular sub-group.

Quasi-experimental approaches are frequently supported by statistical modelling techniques such as probit analysis, survival analysis and hierarchical regression analysis. Probit analysis is designed for situations where linear regression is inappropriate or problematic. Like logistic regression, it can handle dichotomous variables and several different groups of subjects (with abnormal population distribution characteristics) exposed to different levels of stimuli. Survival analysis allows statistical analysis of intervals between two events when the second event does not happen to everyone and when people are observed over different periods of time (Lee, 1980). Another statistical technique often used is hierarchical linear modelling. Hierarchical linear models are useful because, unlike classical analysis of variance models, they do not require that the elements of the within-subjects model be orthogonal to each other (Bryk and Raudenbush, 1992). Perhaps more importantly, as Osgood and Smith point out, hierarchical linear modelling can deliver a powerful test of programme effectiveness with very small samples, because it shifts the unit of analysis from samples of individuals to samples of ‘occasions’, where data are collected on a continuous basis over a significant length of time.

ANNEX II: SELECTED REFERENCES

Bryk, A. and S.W. Raudenbush*, Hierarchical linear models: applications and data analysis methods,* Newbury Park, Sage, 1992

Buckingham, D (2000), *The Making of Citizens*, Routledge, London

Chen, H (2004), *Practical Programme Evaluation*, Sage, London

COPUS (1996), *So did it work? Evaluating public understanding of science events London*, COPUS

Edwards, C (2004), 'Evaluating European Public Awareness of Science Initiatives: A Review of the Literature' in *Science Communication.*2004; 25: 260-271, Sage, London

Guba E and Y S Lincoln (1987), *Fourth Generation Evaluation*, Sage, Beverly Hills

Mackenzie, M and Blamey A (2005) 'The Practice and the Theory: Lessons from the Application of a Theories of Change Approach' in *Evaluation***,** Apr 01, 2005; 11: 151-168

Maienschein, J (1998) 'Science literacy' in *Science* 14 August 1998: Vol. 281. no. 5379, p. 917

Miller, S (2001) 'Public understanding of science at the crossroads' in *Public Understand*. Sci. 10 (2001) 115–120

OECD (2006), Pisa Scientific Literacy Assessment Framework

Office of Science and Technology and the Wellcome Trust (2000), *Science and the Public: A Review of Science Communication and Public Attitudes to Science in Britain London,* OST

Patton M Q (1986), *Utilisation-focused Evaluation*, Sage, Beverly Hills

Pawson R (2006), *Evidence-based policy: a Realistic perspective*, Sage, London

Pawson R and Tilley N (1997) *Realistic Evaluation*, London, Sage

Scriven, M (1997) 'The Theory behind practical evaluation' in *Evaluation*, 1,4, 393-404

Van der Knaap, P (2004), 'Theory-Based Evaluation and Learning: Possibilities and Challenges' in *Evaluation,* Jan 01, 2004; 10: 16-34

Campbell, D.T. and J.C. Stanley, *Experimental and quasi-experimental designs for research*. Chicago, Rand-McNally, 1973

Cook, T.D. and D.T. Campbell, *Quasi experimentation: design and analysis issues for field settings*, Chicago, Rand-McNally, 1979

Finney, D.J. *Probit, Analysis*, Cambridge University Press, 1971

Guba And Lincoln (1989*) Fourth Generation Evaluation*, Sage Publications

Hughes, J, *Philosophy Of Social Research (*1990) Longmans Group Ltd.

Kuhn, T S (1962); *The structure of scientific revolutions*, University of Chicago Press, Chicago

Lee, E.T. *Statistical methods for survival data analysis*, Belmont, 1980

Moberg, D.P., D.L. Piper, J. Wu and R.C. Serlin 'When total randomization is impossible: nested randomised assignment' in *Evaluation Review,* 17:3, 271-91, 1993.

Osgood, D.W. and Smith, G.L. 'Applying hierarchical linear modelling to extended longitudinal evaluations: the Boys Town Follow Up study' in *Evaluation Review*, 19,1, 3-58, 1995

Overall J.E. and C. Klett. *Applied Multivariate Analysis,* New York, McGraw-Hill, 1972

People, Science and Policy Ltd (2005) *Evaluation: Practical Guidelines, London Research Councils UK and Office for Science and Technology*. Available via: http://www.rcuk.ac.uk/aboutrcuk/publications/corporate/evaluationguide.htm

Schwarz, D., R. Flamant and J. Lellouch. *Clinical Trials*, London, Academic Press, 1980

Seltzer, M.H. 'Studying variation in program success: a multilevel modelling approach' in *Evaluation Review,* 18, 3:342-361, 1994

Solomon, P. and J. Draine, 'One-year outcomes of a randomised trial of case management with seriously mentally ill clients leaving jail' in *Evaluation Review*, 19,3, 256-273, 1995

ANNEX III: Glossary

Action research: form of participatory research pioneered by the Tavistock Institute in which researchers are actively engaged 'inside' the organisation or project they are working in. Normally focused on facilitating change and resolving problems generated by change.

Citation analysis: analysis of the frequency and context of 'citations' of specific approaches and methods across a range of indicative information and knowledge sources (MacRoberts, 1989).

Constructivist: so-called 'fourth generation' approach to evaluation and influenced by hermeneutic and methodological approaches. Argues that there is not one single reality but 'multiple realities' associated with the different world views of different 'stakeholders'.

Content analysis: data collection and analysis method that analyses texts.

Control group: the group of subjects in a randomised controlled trial who do not participate in an intervention and whose attitudes or behaviours are compared with those who do participate (the treatment group).

Critical incidents analysis: data collection and analysis technique that retrospectively identifies particular events that were instrumental in shaping an outcome.

Cultural logic: the political, ideological and cultural values embedded in a Science and Society initiative.

Developmental: evaluation that is engaged with the process of the intervention itself and will contribute to changing the way the intervention develops.

Discourse analysis: similar approach to content analysis, but can be extended to include the analysis of the 'sign systems' of whole societies. Also usually adopts a 'critical' perspective that is embedded in a theoretical understanding of social structure.

Ethnographic: generic form of field study methods usually entailing close observation of a study population and the interpretation of their actions.

Evaluation object: the 'thing' an evaluation focuses on.

Evaluation stance: describes the degree of linkage between an evaluation and the intervention being evaluated.

Ex-ante evaluation: assesses an intervention prior to its implementation, usually at the design stage and typically contributing to the formulation of an evaluation plan.

Ex-post evaluation: synonymous with outcome evaluation. Looks at the impacts of an intervention after it has been completed.

Experimentalism: evaluation approach based on empirical verification (demonstrating cause and effect in the real world) and usually involving the use of randomised controlled trials and control-comparison groups.

Focus group: form of group-based interview technique usually involving between six and twelve participants.

Formative: evaluation that looks at the process of an intervention and is typically carried out whilst the intervention is taking place.

Hermeneutic: sub-discipline of interpretative methods. Based on the 'hermeneutic circle' which entails an iterative analysis of a particular phenomenon in order to make the 'hidden meaning' of that phenomenon more clear.

Interpretative: broad social science approach concerned with the study of the products or 'objectifications' of creative, human, historical and cultural activity, typically associated with understanding the meaning of texts.

Intervention: any project, programme or initiative aimed at precipitating change in a target population (e.g. a health promotion campaign).

Knowledge creep: term used to describe the ways in which policy-makers' objectives change over time as they absorb new knowledge.

Logical model: evaluation framework developed initially to support military logistics and subsequently widely distributed by the Kellog Foundation. Typically, divides an evaluation into 'logical' elements of a process, including: goals, purposes, indicators and means of verification.

Life cycle: describes the stages an intervention (and its evaluation) goes through.

Meta-analysis:systematic review of evaluations of Science and Society programmes and projects in order to assess the relevance and potential effectiveness of different evaluation approaches and methodologies.

Modelling: arguably all empirical research involves modelling, but more narrowly, modelling techniques are understood to involve certain forms of engineering or economic analysis, e.g. cybernetic flows of financial transactions. A model is a representation of a system and is used as an 'ideal type' with which to evaluate the effects of an intervention.

Open systems: evaluation approach based on systems theory which considers the organisational and political dynamics affecting an intervention.

Operational evaluation: process evaluation that is directly linked to the management of a project or intervention.

Outcome-oriented: evaluation approach that considers the effects or impacts of an intervention, typically carried out after the intervention has been completed.

Pluralist: evaluation approach combining pragmatic and constructivist ideas with outcome-oriented evaluation.

Pragmatic evaluation: evaluation approach concerned with practical know-how and generally focused on analysing policy.

Process evaluation: regarded as distinct from outcome evaluation and as such considers the how and the why of an intervention rather than whether or not it has an effect.

Randomised controlled trial: experimental technique to evaluate the outcome of an intervention by randomly assigning subjects to two groups: the treatment group and the control group, and then testing whether there are any differences between them that can be attributed to the effects of the intervention.

Social construction: a term derived from hermeneutic and ethnomethodological approaches in social science denoting the existence of multiple realities (i.e. there is no one truth, only different culturally-determined interpretations).

Sociometric mapping: identification of social interaction between actors, for example, communication flows between stakeholders.

Stakeholder: term used in evaluation to denote any actor with an interest in the evaluation and its outputs.

Summative: synonymous with outcome evaluation. Looks at the impacts or 'success' of an intervention and usually takes place after the intervention has been completed.

Task (protocol) analysis: data collection and analysis technique that unpacks human actions (typically jobs) by deconstructing the underlying skill and cognitive components of the action. Usually involves getting subjects to 'think aloud' about why they are performing a particular action.

Theory-driven evaluation: evaluation approach that embodies a core theoretical framework which is used to test whether an intervention has had an impact and in what ways.

Triangulation: 'methodological insurance'. Applying different data collection and analysis techniques to the same evaluation 'object' in order to evaluate that object from a number of different angles.

World view: ideological perspective of a stakeholder; considered to influence the type of evidence that the stakeholder will have a preference for in an evaluation.