MEASURING AND ENSURING EXCELLENCE IN GOVERNMENT SCIENCE AND TECHNOLOGY: INTERNATIONAL PRACTICES

France, Germany, Sweden and the United Kingdom

Report prepared for the S&T Strategy Directorate, Industry Canada
In support of the work of the Council of Science and Technology Advisers

Paul Cunningham
Mark Boden
Steven Glynn
Philip Hills

THE UNIVERSITY of MANCHESTER

PREST
POLICY RESEARCH IN ENGINEERING, SCIENCE & TECHNOLOGY

January 2001

MEASURING AND ENSURING EXCELLENCE IN GOVERNMENT SCIENCE AND TECHNOLOGY: INTERNATIONAL PRACTICES

France, Germany, Sweden and the United Kingdom

Final Report

Report prepared for the S&T Strategy Directorate, Industry Canada
In support of the work of the Council of Science and Technology Advisers

January 2001

Paul Cunningham
Mark Boden
Steven Glynn
Philip Hills

PREST
University of Manchester
United Kingdom

MEASURING AND ENSURING EXCELLENCE IN GOVERNMENT SCIENCE AND TECHNOLOGY: INTERNATIONAL PRACTICES

1 European Study Report: Overview

1.1 Introduction

This report presents the findings of a series of examinations of the approaches employed in France, Germany, Sweden and the United Kingdom which aim to measure and ensure excellence in Government science and technology.

Due to the time and resource limitations of the study it has not been possible to provide a comprehensive account of the approaches used by the component Ministries, agencies and other institutes comprising the national S&T systems in the subject countries for measuring and ensuring excellence. Rather, a range of specific examples drawn from a variety of research actors and fields has been presented for each country, generally supported by a broader overview of the evaluation "scene" found in that country. Whilst the specific examples cannot be regarded as typifying a national approach to assessment, it was nevertheless possible to identify some general trends for each country.

It should also be noted that although the Study Terms of Reference and the SAGE and BEST documents provide a clear indication of the scope of the study, Government S&T in the European context is less easily defined. For example, in the UK many former Government laboratories have undergone partial or full privatisation. Similarly, in the UK and in Germany and Sweden, the complexity of the S&T system has led to the existence of multiple sources of scientific and technological advice and information – all of which may be utilised by Government. For these reasons, the study examined a broader variety of institutions in the expectation that valuable lessons might equally be learned outside of the narrower scope originally defined. The Terms of Reference also emphasise the fact that "government S&T includes both traditional research and development (R&D) activities and related scientific activities (RSA) (i.e., those activities that complement and extend R&D by contributing to the generation, dissemination and application of scientific and technological knowledge)". In the context of this report, 'R&D (or S&T) in support of policy' is understood to include the full spectrum of research activities, from problem oriented studies through the formulation of references and standards, to routine data collection (e.g. for meteorological purposes).

The following section presents the overall summaries from the country studies. This is followed by a brief treatment of some of the broad issues arising, both from the studies themselves, and from the discussions with the CSTA at a presentation meeting in November 2000. Finally, Sections 2 to 5 contain the individual country reports in full.

1.2 Country summaries

1.2.1 France

The last decade has been characterised by a number of organisational and procedural changes in the French science and technology system, including the increasing institutionalisation of evaluation and significant evolution in its domains and methods of application. Evaluation also plays an increasingly important role in policy-making processes.

The French science and technology system can be analysed in terms of four main types of organisations: Ministries with responsibility for Science and Technology; Advisory Bodies; Research Organisations; and the Higher Education Sector.

The Ministry of Research has primary responsibility for research policy, the Ministry of Education has responsibility for the Higher Education system, while a number of other ministries also run research and technology programmes. The formulation of science and technology policy, analysis of progress and the more specific selection of options are tasks delegated to a number of advisory bodies.

The organisation of public sector research in France has undergone significant changes over the past 20 years, particularly in the light of the 1982 Research Act, with budgetary control moved to the Ministry of Research.

There are a number of levels at which evaluation activities may be conducted. At the general level, *ex ante* evaluation is conducted, in terms of the selection of research themes. At the institutional level, of particular interest is the evaluation of research institutions. Evaluation also occurs within research organisations.

At the national level, the structure and operation of two key committees, CNE (Comité national d'évaluation – National Committee for Evaluation) and CNER (Comité national d'évaluation de la recherche – National Committee for the Evaluation of Research), clearly reflect how evaluation has become more important and increasingly institutionalised in France. In addition, the major research organisations all take evaluation seriously and have dedicated units and procedures in place, embracing a full range of activities, from the determination of research themes in line with national policy, through their articulation and impact through research departments, units, teams and projects, to the recruitment and appraisal of individual researchers.

While national policy provides a set of broad objectives and context for activity, research organisations enjoy significant independence, despite being in receipt of government funds. A number also operate under the auspices of more than one ministry, and are relatively free to dispose of their research budget in line with their own internally defined strategic objectives, although these must be coherent with national policy objectives. There must also be a balance between scientific quality and excellence and socio-economic needs. The policy relevance of research and feedback to the policy process are also important features of evaluation.

Institutional evaluation is largely peer based, although the value of quantitative approaches is recognised, and in cases peer review is being supplemented by use of indicators. Whilst evaluation is becoming much more systematised, it requires informal and ad hoc dimensions to retain flexibility.

The main weaknesses of the evaluation systems in French organisations are the bureaucratic procedures and the work entailed by the numerous committees, peer reviewers and so forth.

1.2.2 Germany

The most striking feature of the German S&T system is the diversity of research actors of which it is composed. These include: universities and other higher education institutions (conducting mainly basic/long-term application oriented research); the "national research

centres" of the Helmholtz Society (mainly strategic research); research institutes of the Max Planck Society (mostly basic research); the Leibnitz or former 'Blue List' institutes (generally carrying out medium-term pre-competitive and "service" research; and the Fraunhofer Society institutes (mostly performing applied/application oriented research). In addition there are a number of Government research institutes (which undertake "departmental" research in support of ministerial objectives) and industry (which includes the research activities of the Confederation of Industrial Research Association together with its own R&D activities).

This diverse research base presents the Government (which consists both of the Federal *Bund* and the State or *Länder* governments) with a range of sources from which it may commission research and obtain advice, and they are not specifically reliant on their 'in-house' agencies. Moreover, the government(s) are not the sole customer for such research and institutions offer their services to a broad range of clients. Further complexity is added to the system through the frequent formation of research consortia involving institutes from the same or other sectors. This complexity is not necessarily a negative point and it is argued that the complexity of policy questions requires a wide network of researchers. Despite the broad oversight role exercised by the Federal Ministry of Education and Research (BMBF), there is no strong overall centralised policy body.

The German evaluation system tends to reflect the complexity of the S&T system. Evaluation (broadly defined) is operated at a number of levels: the system level, the institutional level, at the programme level (often through the use of "impact analyses"), and at the group or individual level. There is also much use of *ex ante* evaluation approaches and technology assessments. Many evaluations, particularly at the system and institutional levels, are undertaken at the instigation of the BMBF and the Science Council (Wissenschaftsrat) often plays a significant role in evaluation activities. As in many other countries, peer review forms the most common method for ensuring excellence, although care is taken to include a wide yet relevant representation on panels (including industry and other users of the research).

The German study focused on the following bodies: the Ministry for Economics and Technology (BMWi), including its Institute for Materials Research and Testing (BAM); the Ministry for Food, Agriculture and Forestry (BML); the Ministry of Education and Research (BMBF), with a focus on two of its internal programmes; the Federal Ministry for Health (BMG); and the National Research Centres (Helmholtz Society).

As in other European countries, it was not possible to define a national system for measuring and ensuring excellence in S&T, although a number of general trends and characteristics were identified.

The programme definition process is very consensus-oriented, with extensive use of external scientific and technical advice, together with inputs from the eventual end-users of the research. The latter generally include industry, but the broader public may also play a role. Project definition is also highly influenced by external inputs.

The planning process, particularly at the project level, is strongly emphasised and involves a clear definition of milestones and objectives. The eventual use of results and plans for their dissemination may also be taken into account. This aspect is, in some ways, a prerequisite for the next feature.

Project management is frequently undertaken by third parties. These are often from specialist external management organisations or 'Projekttrager". Such managerial tasks range from relatively simple administrative and financial management to monitoring and evaluation. In addition, much emphasis is placed on project monitoring. This may be conducted through a reporting system which typically will involve the preparation of mid-term and final project reviews.

The scientific and technical capability and capacity of research performers are assessed generally either through open tendering and proposal processes which scrutinise the feasibility of intended projects and the ability of research teams to undertake them (possibly using peer review); and through systemic evaluations of individual institutions or groups of institutions.

Although in some cases there is systematic evaluation of immediate research impacts, which is generally achieved by requesting the relevant information to be provided in the final project reports, the emphasis on final reports tends to neglect any longer-term follow-up procedures aimed at identifying downstream impacts and effects.

1.2.3 Sweden

The Swedish system of Government differs from that found in many other countries. The ten Ministries (plus the Prime Minister's Office) which comprise the Government Office are very small, but are supported in their executive functions by a number of National Boards and Agencies. These are independent of the Ministries, although they may report to them. A number are also autonomous as regards regulation in their specific areas. For example, the Environment Protection Agency (EPA) is, in staff terms, about 5 times the size of the Environment Ministry and holds the main core of expertise. Although the EPA is responsible to the Ministry, and receives (very) broad instructions as to how it should allocate its funds, it is a separate body. On January 1st 2001, the (basic research oriented) research councils under the Ministry of Education and Science, together with a number of more sector specific (applied research oriented) research councils were integrated into one large Science Council. At the time of writing, it was not fully clear as to how the new system will operate. The Swedish system is further complicated by the existence of private research foundations. These private foundations can be seen as mirroring the research council system but in the private sphere (although the government appoints their boards). In fact, the foundations actually disburse more money than the research councils do.

Given this rather diffuse (and changing) nature of research in Sweden it is difficult to assess where the 'government' is involved in evaluating research and ensuring excellence. Thus, rather than a centralised system, responsibility for these activities appears to be largely devolved to the individual organisations concerned with performing research.

The study looked at the following agencies, foundations and research councils: The Swedish Environmental Protection Agency; Swedish Foundation for Strategic Environmental Research (MISTRA); Natural Science Research Council; Swedish Foundation for Strategic Research (SSF).

Overall, the Swedish system was characterised by a highly devolved approach to measuring and ensuring excellence in that ministerial direction over programme definition was very broad; in all the examples reviewed, quality assessment was undertaken at the agency, council or foundation level. No centralised guidelines were apparent for ensuring excellence

across the entire Swedish S&T system. However, a number of common themes were identified. The relevance of the S&T research undertaken was an important consideration in the agencies and foundations, whilst the research councils placed emphasis on scientific quality. With regard to evaluation (in its broadest sense), Sweden, in common with many other countries, relied heavily on peer review, particularly for proposal assessment. As a small country, there was a strong reliance on the use of international experts for quality assessment. Attention was also focused on the importance of disseminating the results of research, both to an academic audience and to the broader public. Lastly, there was evidence to suggest that the timeliness of the outputs of scientific research was a very important consideration – a feature which appeared to be facilitated by close communications between the sponsors and users of scientific and technological research.

1.2.4 The United Kingdom

This UK report focuses first on those central departments (Treasury, Cabinet Office and Office of Science and Technology) that provide a general framework for S&T activities and then on the 'executive' departments that act within that framework. The central departments provide guidance and advice on issues of broad principle, priorities and strategy. They exercise influence mainly through the control of funds (in the case of the Treasury) and by issuing advice and guidance. They also have some coordinating functions. The Cabinet Office, in particular, have recently taken significant steps towards improving the policy making machinery.

The detailed management of science and technology is devolved to other departments and the paper summarises the relevant activities of those responsible for Trade and Industry, Transport, Environment and the Regions, Agriculture, Fisheries and Food, and of Health. It also touches on some of the agencies, such as the Food Standards Agency and the National Physical Laboratory (now under private management).

The distinctive characteristics of the UK system are its devolved nature, its emphasis on a culture that focuses on well defined objectives and (usually) *ex post* evaluation and on peer review. There is an evolving trend towards eclecticism (enlisting a wider range of scientific opinion through open consultative processes) and greater transparency (by contrast to a rather secretive tradition that has been found wanting). The now widely accepted ROAME system and its variants have been adapted to the differing needs of departments and is now applied with considerable flexibility. This approach has had a significant impact on the science management culture.

The paper describes the specific arrangements within most of the departments with major science and technology responsibilities. The DTI section provides some detail on evaluation methodologies which are to a considerable degree typical of departments generally.

The whole is drawn together in a final section that attempts to identify the main prerequisites and impediments to excellence as exemplified in the UK. It also suggests approaches that have given rise to problems largely by reference to those that have replaced them.

1.3 Broad issues/findings

This section deals with a number of issues which are common to several, if not all, of the country reports and also addresses a number of specific points raised by the CSTA.

1.3.1 "Advice" versus research

Perhaps the major finding from the European studies is that, in contrast to the 'North American' models, Government S&T appears to be performed by a broader research base which extends beyond the Government sector. Thus, Government is less the manager and more the customer for the research. Moreover, there is an increasing tendency for European Governments to seek scientific advice (as a "package") rather than initiating specific research to answer their questions. That is, Government "taps in" to existing scientific expertise from a variety of directions and at a range of levels. If this advice is not able to provide adequate policy guidance then Governments must decide whether there is a need to define specific programmes of research in order to obtain the answers they seek.

1.3.2 Outsourcing and privatisation

The increasing outsourcing of S&T, which includes the privatisation of Government laboratories witnessed in the UK in particular, can be a double-edged sword, however. Government is no longer able to retain direct control over the sources of its scientific advice (for example, to ensure excellence or to ensure that research expertise is maintained in critical areas), but these sources have become broader, with a greater overall and more comprehensive research capacity. Government is thus forced to ensure that it maintains the health of this broader science base rather than fostering its in-house facilities alone, although the task is alleviated by the fact that Government becomes but one source of funds amongst several. In France, where the system of centralised Government control is becoming more relaxed, although still a far cry from the ethos of privatisation espoused by the UK Government, the greater autonomy conferred upon research agencies and researchers brings with it an attendant requirement for greater accountability, transparency and evaluation.

The issue of privatisation and its effects on former Government laboratories is moot, but beyond the resources of this particular study. No specific evidence was collected regarding the impact of privatisation on ensuring excellence.

1.3.3 Prerequisites for excellence and "best practice"

In terms of defining the "prerequisites for excellence", in which the CSTA was particularly interested, the two salient factors which emerge from these studies appear to be a) ensuring the timeliness of the research – i.e. ensuring that the results are available sufficiently early to be able to influence policy, and b) ensuring the existence of a capacity to perform necessary research and that this is accessible to policy makers. In a sense, the first can be achieved by linking policy making closely to scientific advice – advanced notice of an emerging research need allows better preparation than responding to an emergent problem. The UK report, in particular, presents a number of mechanisms by which this goal is attempted.

At the operational level, it could be argued that a clear prerequisite for excellence is to have in place systematic, yet relatively flexible evaluation processes. The absence of such a systematic approach, *ipso facto*, would prove an impediment to excellence. Within the limited scope of this study, it has not been possible to identify, yet alone benchmark, all the possible approaches to measuring and ensuring excellence in S&T. However, two possible candidates might be suggested: the ROAME approach employed by the UK Department of Trade and Industry and adopted, in modified form, by a number of other UK Ministries; and the in-depth planning and monitoring approaches adopted by several of the German examples.

The ROAME process is clearly applicable throughout the continuum for S&T, whilst the German approaches are less concerned at the end-of-project/programme evaluation stage. However, in respect of the need for timeliness in Government S&T, evaluation of the results of the research may be less important than evaluation of the process of research (its effectiveness and efficiency) and the capacity of the research performers. Indeed, the feedback element of ROAME (sometimes designated by a additional "F" to the acronym) aims at incrementally improving the overall research process from planning to output. Lessons learned from the application of the ROAME process are dealt with in more detail in the UK report.

1.3.4 Evaluation burden

It is clear that evaluation (from *ex ante* review to studies of results and impacts conducted *ex post*) carries a cost in both time and resources. The need to strike a balance between the benefits of evaluation and its attendant costs have long been a dilemma to policy makers. The New Zealand and Australia report mooted the idea of classifying S&T into three categories based on the level of complexity and use of the science in order to determine the level of evaluation required. This is an interesting concept which does not seem to have found immediate realisation within the systems studied for this report. However, that is not to say that agencies do not adapt the intensity of their evaluation according to the "significance" of the research they commission or undertake. In the UK, it is known that Government agencies may adopt a general 'rule of thumb' with regard to the resources they allocate to evaluation – generally this falls somewhere between 0.5% to 1.0% of the total programme budget. However, not all programmes or projects may actually be subject to an evaluation.

Related to this issue is the idea of "peer review fatigue". This appears to be an emergent issue, certainly in both France and the UK, caused by the extremely widespread use of this mechanism in project/programme definition, proposal selection and in *ex post* evaluation. According to some of those interviewed in the UK, the phenomenon is manifested through a reticence amongst the scientific community to take part in peer review activities and, in one instance, to a perceived deterioration of the quality of peer review. One obvious response is to utilise peer reviewers from outside the country (as is the case in France), an approach long adopted by smaller countries where the search for independent viewpoints has been hampered by the small size of the national research community. Another, rather interesting approach, has recently been adopted by the UK Engineering and Physical Sciences Research Council (EPSRC). The research council is reported to be considering a scheme whereby peer reviewers of project proposals may soon be able to collect 'peer miles' as part of a reward scheme. The scheme in question would award points to individual academics for submitting reviews on time, which would add to a total that could be redeemed by their department at the end of each year. The EPSRC expect that any resources received by departments as part of such a scheme would be used for research purposes, for instance to support researchers' or research students' conference expenses. The Council has been increasing its emphasis on the funding of unsolicited, peer reviewed research proposals – the so-called "responsive" mode of funding – and is anxious to reform the peer review process to give greater credit to adventurous or interdisciplinary work A reward system for reviewers is seen as a possible means of increasing the referee base available to it.

1.3.5 Transparency

The reports note a number of approaches used by agencies to ensure transparency and effective communication of their S&T activities. Transparency of Government S&T issues, advice and policy decisions have formed a major concern over recent months in the UK – not

least because of the BSE furore and the debate over GM crops. It is also possible to identify specific examples at other levels, for example the publication of research strategy consultation documents by departments such as the Ministry of Agriculture, Food and Fisheries. Another example concerns the Department of Health whose researchers are encouraged to make their work widely available, using a number of dissemination routes, including a National Research Register (set up with NHS R&D funding) and the department's website. Very recently, the UK Biotechnology and Biological Research Council (BBSRC) has launched a new web page[1] to which members of the public, other scientists and special interest groups can communicate their comments regarding the Council's areas of research.

In Sweden, the requirement for researchers to publish 'public documents' was also noted, and in several countries, the publication, on the Internet, of the general results of ministerial research activities appears to be growing.

1.3.6 Ethical considerations

The CSTA asked for input regarding the handling of "ethics" in the context of excellence. This is certainly an emerging issue in the field of R&D evaluation. For example, Rip (in prep.)[2] acknowledges the pressures imposed on the process of R&D evaluation through the need to address the new stakeholders for research. He notes the emergence of a new acronym - ELSI - generated by the growing interest in Ethical, Legal and Social Implications of R&D.

No specific evidence was obtained on this issue in the studies, apart from a general tendency noted throughout for greater public involvement in S&T policy setting and in the assessment of research outputs. In addition, ethical advice is beginning to play a role alongside scientific advice – indeed, few contemporary scientific questions are free from an ethical dimension – witness the establishment of the UK's Human Genetics Advisory Council, a high level policy advisory body. At the departmental level, for example, in the UK some of the Department of Health's scientific advisory committees, which advise on new scientific developments and present recommendations for further work, have held open meetings and involved members of interest groups and consumer organisations.

[1] www.bbsrc.ac.uk/society/consult

[2] Rip, A. Challenges for R&D evaluation. (in prep.) – based on a presentation at the USA-European Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany, 11-14 September 2000.

2 European Study Report: France

Prepared by Mark Boden

2.1 Introduction

This report first presents a brief description of the French research system to provide a clearer picture of the context within evaluation, the term taken here to connote the full spectrum of evaluation and assessment activities, takes place (Section 2.2.1). This is followed by a general overview of the French evaluation scene (Section 2.2.2). A range of specific examples drawn from a variety of research actors and fields is then presented (Section 2.3). These show both overarching approaches to evaluation and the approaches of particular mission oriented research organisations, which are towards the leading edge of evaluation practice. The next section (Section 2.4) considers what general trends can be identified.

As Philippe Laredo points out in a recent OECD paper on Evaluation in France[3], the last decade has been characterised by a number of organisational and procedural changes in the French science and technology system. Against this background, evaluation has become increasingly institutionalised, with significant evolution in its domains and methods of application still ongoing. He notes two key features of evaluation in France: its recent development and its locus outside of the organisations traditionally responsible for monitoring government expenditure.

Laredo also notes the increasingly important role of evaluation in policy-making processes, which he sees as founded on a rich and longstanding tradition of advisory bodies and the use of quantitative indicators of scientific and technological progress. A further related trend he observes is that of the role of research institutions in directing research and determining its directions, mediating between research and policy activities, rather than merely conducting research. For him the challenge is to achieve an appropriate balance between the scientific credibility of research activities and their relevance to the policy process.

2.2 Main Findings

2.2.1 Overview of system

The French science and technology system can be analysed in terms of four main types of organisations: Ministries with Responsibility for Science and Technology; Advisory Bodies; Research Organisations; and the Higher Education Sector.

2.2.1.1 Ministries

The Ministry of Research[4] (MR) has primary responsibility for research policy. It advises public and private research organisations (such as the CNRS) and proposes specific measures and financing by the Government. Within this ministry, the Technology Unit is responsible for enhancing the value of public research and co-operation between enterprises and public research. This unit is also responsible for various technological organisations, including: the National Environmental Agency (ADEME), the Agency for Technologies Diffusion (ADIT),

[3] Laredo, P. (1997) Evaluation in France: a Decade of Experience in OECD *Policy Evaluation in Innovation and Technology: Towards Best Practices*. See also: Mustar, P. and Laredo, P. (forthcoming) Innovation and Research Policy in France (1980-2000) *Research Policy*.

[4] www.recherche.gouv.fr

the National Agency for the 'valorisation' of research (ANVAR) - under the joint authority of the Ministry of Economy, Finance and Industry, the Atomic Energy Commission (CEA), the French Research Institute for Exploitation of the Sea (IFREMER), and the Agricultural and Environmental Engineering Research Institute (CEMAGREF). Another unit of the Ministry, the "Direction de la Recherche" evaluates research programmes launched by the Ministry, and is responsible for the co-ordination of public research organisations under the supervision of the Ministry of Research.

The Ministry of Education[5] (MEN), has responsibility for the Higher Education system, and its associated research institutes, relating to educational matters. Independent from, although formally attached to the MEN and its responsibility for the higher education system, is the Comité national d'évaluation (CNE). The CNE examines and evaluates all aspects of university activity and that of the research organisations under the MEN's area of responsibility in France.

A number of other ministries also run research and technology programmes, which, in many cases, are co-financed by the Ministry for Research.

Within the Ministry of Agriculture and Fisheries[6] (*Ministère de l'Agriculture et de la Pêche*) research policy is coordinated by the DGER, which also has joint responsibility with the Ministry for Research for the following research organisations: INRA, CEMAGREF, CNEVA, and IFREMER (see below).

The Ministry of Equipment, Transport and Housing[7] (*Ministère de l'équipement, des transports et du logement*) has responsibility for a scientific and technical equipment network, comprising a number of engineering schools and services. Its activities are centred on the development of skills and include research. In particular, in cooperation with other ministries and professionals in the relevant sectors, a number of collaborative research programmes (bringing together public and private sector researchers) are in operation, namely:

- PREDIT, which supports innovation in the field of terrestrial transport;
- RGC&U, a programme of collaborative R&D in civil engineering; and
- PUCA, an interministerial programme of research in the field of urbanisation, architecture and construction.

The Ministry of the Environment[8] (*Ministère de l'Aménagement du Territoire et de l'Environnement*) finances and manages research programmes in a number of environmental areas. Their general aim, in support of government policy, is to better manage the natural environment and resources and to reduce pollution. The normal procedure, following a call for tender and the review and selection of project proposals by an orientation committee and a scientific council, is to let research contracts under these programmes to both public and private research organisations.

[5] www.education.gouv.fr
[6] www.agriculture.gouv.fr
[7] www.equipement.gouv.fr
[8] www.environnement.gouv.fr

2.2.1.2 Advisory Bodies

The formulation of science and technology policy, analysis of progress and the more specific selection of options are tasks delegated to a number of advisory bodies, operating both within the government and deliberately independent to it. These bodies also play a role in the evaluation of policy programmes and ensuring that such results take their place on the policy agenda.

The Parliamentary Office for Scientific and Technical Options[9] (OPECST) comprises members from both houses of parliament. Issues are referred to OPECST by the board of either assembly (at the request of the chairman of a political group, or on the initiative of sixty Members of the National Assembly or forty Senators), or by a special or permanent committee. The issues considered to date fall into four main categories: energy, environment, new technologies and life sciences. OPECST also operates a Scientific Committee, which plays an intermediary role between the domains of policy and of research. The membership of this committee reflects the full range of scientific and technological disciplines, and is composed of fifteen leading figures selected for their expertise.

The Inter-Ministry Committee of Scientific and Technical Research (CIRST) reports directly to the Prime Minister. It meets once a year to set up the national orientations of the research and technology policy and is formally attached to the Ministry of Research.

On behalf of CIRST, the ***Science Academy*** produces a bi-annual report (RST)[10] on the situation of research in France. Its conclusions provide the government with an analysis of the activities of several research organisations, their international status and, on this basis, a vision of the future. The Academy has set up a special committee (Comité RST) to carry out this study, comprising members of the Academy's administrative bureau and its membership, particularly members of its Applications Council (CADAS).

The ***National Council of Science*** reports to the Minister of Research, and comprises 20 public and private national and European researchers. They advise the Government on research policy.

The ***Commissariat General du Plan*** is a commission made up of industrialists, trade unions and representatives from various public organisations. This group regularly proposes long-term policy to the national government. Their proposals have led to the establishment of large programmes in space, nuclear science and aeronautics.

The ***Comite d'Orientation Strategique*** (COS), is composed of 15 members nominated by the Prime Minister, was created in 1995. The priority task of this committee is to prepare the annual report on the national research strategy to be presented by the Minister to Parliament.

The ***Observatoire des Sciences et des Techniques***[11] (OST), is an independent body funded by several ministries, in charge of the production of French S&T indicators. It was created at the same time as the CNER (see below) as an integral part of the evaluation system.

---

[9] www.assemblee-nat.fr
[10] Available from: www.academie-sciences.fr
[11] www.obs-ost.fr/ost_fr

There are also two key committees, with important advisory roles, concerned with the evaluation of research, CNE and CNER. These are discussed in more detail in the next section.

2.2.1.3 Research Organisations[12]

The organisation of public sector research in France has undergone significant changes over the past 20 years, particularly in the light of the 1982 Research Act. Most government research laboratories essentially became part of the French civil service, being accorded the official status of "EPST", public scientific and technological establishment (*Établissement Public à Caractère Scientifique et Technologique*). Under the act, the budgetary control of EPST's was moved away from their parent ministries to the Ministry of Research, hence the joint responsibility for a number of research institutes (as noted above). The overlaps and links between government research institutes, university research, and company research are also increasingly complex[13]. There are 9 EPSTs currently in existence. This category of institute is of particular interest to this study and contact was made with a number of the following:

CEMAGREF (www.cemagref.fr - Centre national du machinisme agricole, du génie rural, des eaux et des forêts) conducts agricultural and environmental engineering research. Recent work has centred on sustainable development in non-urban areas, and relies on an interdisciplinary approach. In addition to fundamental research activities, CEMAGREF provides services to public and private enterprise. The four main areas of interest are: hydrosystem functioning; water and waste disposal engineering and management; management of predominantly rural areas; and agricultural and food engineering

CNRS (www.cnrs.fr - Centre national de la recherche scientifique) The National Centre for Scientific Research. The CNRS is a large public basic research organisation. It comprises 1,300 service and research units spread throughout France and covering all fields of research : Physical Sciences and Mathematics, Nuclear and Particle Physics, Sciences of the Universe, Engineering Sciences, Chemical Sciences, Life Sciences, Humanities and Social Sciences

The National Committee of Scientific Research of the CNRS evaluates researchers and laboratories belonging to CNRS. It was restructured in 1995. The evaluation process is meant to guarantee the quality of the research and give CNRS international legitimacy. The National Committee also has an important role in the recruitment of researchers.

INED (www.ined.fr- Institut national d'études démographiques). The National Institute for Demographic studies conducts research, on various aspects of populations.

INRA (www.inra.fr- Institut national de la recherche agronomique) The National Institute of Agricultural Research currently has three principal objectives:

- To guarantee consumers high-quality food;
- To ensure that agricultural and agro-food companies are competitive; and

[12] Ministry of Education *Major French Research Institutions.*

[13] Chastgnaret, G. and Goujon, M. (1998) *Propositions pour simplifier, deconcentrer et mutualiser les functions de gestion de la recherché dans les Etablissements Publics Scientifiques et de Technologie* report to the Education Minister.

- To contribute to integrated land development and sustainable management of natural resources.

INRA's evaluation activities are considered in more detail in the next section.

INRETS (www.inrets.fr - Institut national de recherche sur les transports et leur sécurité). The French National Institute for Transport and Safety Research is a state-financed scientific and technological body under the dual administrative supervision of the Ministry of Research and the Ministry of Transport. It is responsible for the organisation, conduct and assessment of technological research and development concerned with the improvement of the means and systems of transport and of traffic from technical, economic and social viewpoints. It also conducts evaluative and advisory studies within these domains and promotes research results.

INRIA (www.inria.fr - Institut national de recherche en informatique et en automatique).The French National Institute for Research in Computer Science and Control is the joint responsibility of the Ministries of Research and Industry

INSERM (www.inserm.fr - Institut national de la santé et de la recherche médicale) The French Institute of Health and Medical Research is overseen jointly by the Ministries of Research and Health, and has the primary aim of promoting health, through improved understanding of diseases. INSERM's evaluation activities are considered in more detail in the next section.

IRD (www.ird.fr - Institut de recherche pour le développement). The IRD is the joint responsibility of the Ministries of Research and Foreign Affairs. Its scientific research is centred on tropical and Mediterranean environments. Consideration of its evaluation practices is given in the next section.

LPCP (www.lcpc.fr/- Laboratoire central des ponts et chausses) works both for the State and local authorities in the fileds of civil engineering, transport, urban engineering and environment.

The allocation of state funding to "EPSTs" in 1999, a total annual budget of around 20 billion Francs, was divided as follows:

CNRS	61%
INRA	16%
INSERM	12%
IRD	5%
Others	4%
INRIA	2%

In addition to the ESPTs, other public sector research organisations were classified according to the character of their activity, as follows:

Etablissements publics à caractère industriel et commercial (EPIC). The following agencies interact to a much greater extent with industrial actors, being more mission oriented in their activities than ESPTs.

ADEME (www.ademe.fr- Agence de l'environnement et de la maîtrise de l'énergie) is the Agency for Environment and Energy Management and under the joint tutelage of the Ministries of the Environment, of Industry and of Research. However, it does not have its own research facilities, but finances and manages research conducted by public research bodies, universities, companies and technical centres, which is oriented towards practices and technologies in support of policies for energy and environmental management.

ADIT (www.adit.fr/ - Agence pour la diffusion de l'information technologique). The Information Technology Diffusion Agency is jointly under the Ministries of Research and Foreign Affairs. As with ADEME its concern is with the promotion of research and the diffusion of research results in support of industrial development.

ANDRA (www.andra.fr/- Agence nationale de gestion des déchets radioactifs). The National Agency for Radioactive Waste Management is jointly under the Ministries of Environment, Industry and the Environment, and, as its name suggests, is concerned with improving the management of radioactive waste for the long term protection of the environment. Its main research effort centres on the disposal of wastes (HLWs) in deep geological formations, and the agency runs a number of dedicated underground research laboratories.

ANVAR (www.anvar.fr - Agence nationale de valorisation de la recherche). The National Innovation Agency is the joint responsibility of the Ministries of Research and Industry. Its main activity is the provision of finance and advice to SMEs, laboratories and entrepreneurs regarding innovation and related financial and commercial issues.

BRGM (www.brgm.fr - Bureau de recherches géologiques et minières). The Bureau of Geological and Mining Research is under the dual tutelage of the Ministries of Research and Industry, conducting, coordinating and funding research and providing expertise and advice in the sustainable management of natural resources in surface and subsurface domains.

CEA (www.cea.fr - Commissariat à l'énergie atomique) The Atomic Energy Commission, under the Ministries Research, Defence and Industry, conducts research in support of government policy for defence, nuclear and alternative energies. Its evaluation activities are considered in the next section.

CIRAD (www.cirad.fr - Centre de coopération international en recherche agronomique). The Centre for International Co-operation in Agronomic Research reports to the Ministries of Research and Foreign Affairs and specialises in agronomic research in hot regions, with network of overseas laboratories. It also specialises in the provision of advice and other services and in promoting technology transfer.

CNDP (www.cndp.fr - Centre national de documentation pédagogique). The National Centre for Teaching and Learning Documentation is a national network of documentation centres under the Ministry of Education.

CNED (www.cned.fr - Centre national d'enseignement à distance). The National Centre for Distance Learning is also under the Ministry of Education is concerned with the development and promotion of distance learning in the francophone world.

CNES (www.cnes.fr - Centre national d'études spatiales). The National Centre for Space Studies, under the Ministry of Research, is charged with the elaboration and implementation

of French space policy. It acts in partnerships with research actors in both industry and across the public sector.

CSI (www.cite-sciences.fr/ - Cité des sciences et de l'industrie). The Science and Industry City is a national centre for scientific and industrial communication.

CSTB (www.cstb.fr - Centre scientifique et technique du bâtiment). The Centre for Building Science and Technology (Ministère de l'Équipement, du Logement et des Transports) is a centre of expertise in construction science and technology, conducting research, and providing expert consultancy and advice.

IFREMER (www.ifremer.fr/francais/ - Institut français de recherche pour l'exploitation de la mer). The French institute for research and exploitation of the sea), is under the joint authority of the Ministries of Education, Research and Technology, Fisheries, and Amenities, Transport and Housing.

INERIS (www.ineris.fr - Institut national de l'environnement industriel et des risques) is under the supervision of the Ministry for national and regional development and the environment, and is concerned with environmental safety. Its mission centres on the assessment and prevention of accidental and chronic risks to people and the environment due to industrial plants, chemical substances and underground operations, and its activities include research and the provision of expertise in response to public authorities and industrial operators.

ONERA (www.onera.fr - Office national d'études et de recherches aérospatiales). The French Aeronautics and Space Research Centre reports to the French Ministry of Defense. The expertise of ONERA covers all the range of scientific disciplines involved in aircraft, spacecraft and missile design. Its activities range form basic research through to technology transfer and commercialisation activities and the support of government civil and military aerospace policy.

Etablissements publics à caractère administratif (EPA) are a further category of government affiliated organisations, with an administrative dimension to their activities, but which also conduct related research. These include:

CEE (www.cee-recherche.fr - Centre d'études de l'emploi). The Centre for Employment Studies is attached to the CNRS and reports to the Ministries of Research and Employment.

INRP (www.inrp.fr - Institut national de recherche pédagogique). The National Institue for Teaching and Learning Research reports to the Ministries of Research and Education.

Other Research Organisations

Research is conducted by a number of private foundations of recognised public utility including: the CEPH (Centre d'étude du polymorphisme humain), the Institut Curie (www.curie.fr) and the Institut Pasteur (www.pasteur.fr).

Groupements d'intérêt public (GIP) are also involved in the promotion, conduct, finance and evaluation of research. These include: the National Agency for AIDS Research (ANRS – www.anrs.cnrs-mrs.fr); the National Genotyping Centre (CNG - www.cng.fr); the National Gene Sequencing Centre (CNS - www.genoscope.cns.fr); the OST (Observatoire des sciences

et techniques - discussed above); and RENATER (Réseau national pour la technologie, l'enseignement et la recherché - www.renater.fr).

2.2.1.4 The Higher Education Sector.

As with the above institutions, in 1984 the legal status and organisational structure of the universities, and other higher education institutes was changed. They became *Etablissements Publics de Caractere Scientifique, Culturel et Professionnel* (EPSCP). These are further subdivided into:

- 82 Universities[14]
- National Polytechnic Institutes
- 3 Ecoles Normales Supérieures
- 14 Large Establishments of varied status[15]

2.2.1.5 Federal Research Institutes (IFR)

Over recent years common interests among laboratories belonging to different research organisations, of varied official status, particularly in the life sciences, led to the establishment, in 1994, of a series of Federal Research Institutes, under the auspices of the Ministries of Research, Education and health and their attached research organisations (CNRS, INSERM, INGRA, IRD, CEA, the Conference of University Presidents). They support the implementation of a visibly coherent science policy, and allow an optimal use of intellectual and physical resources and the development of relations. Their operation is overseen by both a steering committee and a scientific council. The former, which is composed of representatives of the above partner organisations, takes responsibility for their strategic direction and the allocation of funds on the basis of evaluations conducted by the latter, which is composed of scientific experts selected by the partner organisations. Evaluation of potential projects by the experts is based on the use of *ad hoc* evaluation grids. The renewal of existing projects entails a site visit by a group of three experts. The criteria applied in both instances include the relevance of the project, the means available, national and international strategic interest, scientific objectives, the value added provided by the IFR, its feasibility, its socio-economic and commercial potential, and its training benefits.

2.2.2 Overview of evaluation

There are a number of levels at which evaluation activities may be conducted.

2.2.2.1 General level

The general level of evaluation is *ex ante* evaluation in terms of the selection of research themes. The various advisory bodies described above include assessments of future research directions amongst their policy recommendations. The CNE and CNER (detailed below) while focusing their activities on the institutional level also operate at the more general multi-organisational level.

---

[14] For a complete list see www.education.gouv.fr/sup/univ.htm

[15] Ecole centrales in Lille, Lyon and Nantes, the Ecole nationale des Ponts et Chausées, the Ecole nationale supérieure des arts et industries de Strasbourg, the Instituts nationaux des sciences appliquées in Lyon, Rennes, Toulouse and Rouen, the Institut supérieur des matériaux et de la construction mécanique, and the Universités de technologie in Compiègne, Belfort-Montbéliard and Troyes.

2.2.2.2 Institutional Level

At the institutional level, of particular interest is the evaluation of research institutions. Again, this type of evaluation is conducted on behalf of the state. With the joint tutelage system of many research organisations, this arrangement ensures a degree of independence, while also maintaining government involvement. However, at the general (i.e. central government) level in France, there are two main bodies concerned with the systematic evaluation (in terms of both assessing performance and determining future directions) of scientific and technological activity in institutions. These are: the National Committee for Evaluation (CNE), which is concerned with the evaluation of higher education institutions, and the National Committee for Evaluation of Research (CNER), which covers the rest of French public sector research. Their structures, roles and operating methods are described in detail below. In addition within the CNRS, there is an audit section.

The eventual aim of the two bodies is to comprehensively evaluate all institutions, but to date they have been selected on more of an *ad hoc* basis. There are, according to Laredo[16], two key dimensions to evaluation activities at this level:

- the robustness and credibility of the evaluation; and
- the embedding of strategic evaluation in the evaluated bodies

2.2.2.3 Internal Level

Evaluation also occurs within the types of research organisations described above. There has been a trend towards growing autonomy, with organisations steering and conducting their own activities, including the evaluation of research themes, of their constituent units (especially theme-oriented units), and projects (see INRA/INSERM examples). This greater autonomy is related to the change in the ministerial affiliation and control of research organisations. This type of evaluation tends to be more institutionalised and systematic in nature. As the specific cases of INSERM and INRA in the next section illustrate, this level is the locus of debate over evaluation practice, and the appropriateness of methods to institutional types. There is currently greater consideration being given to the use of quantitative approaches in assessing internal performance, such as bibliometric tools. At a deeper level within these organisations, evaluation also occurs within research units, focusing on the evaluation of personnel and project progress. This tends to employ common methods and practices throughout the parent research organisation.

2.3 Specific cases

This section concentrates on the evaluation activities of a selection of organisations operating at different levels in the French system. Because of time and resource limitations, a comprehensive, in-depth view of the entire national system for the assessment of research has not been possible. A number of specific case studies of ministries and research institutes are presented below.

Of particular interest is a longer-term study currently being conducted by the CNER, which is taking a much more comprehensive view of the evaluation issues and practices of French research organisations.

[16] Laredo *op. cit.*

2.3.1 CNE - Comité national d'évaluation[17]

Since its establishment (by decree) in 1989, the principal mission of the CNE is the evaluation of the universities, colleges and the large establishments, which report to the Ministry of Education, as well as those attached to other ministries. The CNE reports directly to the French President, who selects its 17 members (11 of whom are drawn from the academic and scientific communities, 4 from the Conseil économique et social, and one each from the Conseil d'État and the Cour des Comptes). It is assisted by its own secretariat general. The CNE's evaluations cover the training, research, and technology transfer activities of these institutions, as well as aspects of their policy and management. However, their evaluation remit does not extend to individuals or to the distribution of state funding. In conducting an evaluation the CNE provides a series of conclusions and recommendations. Although these are not legally binding, the extent to which they have been adopted and their impacts are followed up 18 months after the publication of the CNE's report.

The CNE determines its own working methods and selects the organisations for scrutiny. In carrying out an evaluation of an institution, the initial decision for the evaluation to take place is determined by the CNE itself. It has already evaluated all the French universities and most of the Ecoles, and has embarked on a second wave of return evaluations to many institutions. In additional to evaluations of individual organisations, the CNE conducts multi-organisational evaluations, examining the interactions between organisations operating in a specific geographical area, as well as "transversal" evaluations of specific themes in higher education. On the basis of its various evaluation studies, the CNE also produces an annual report for the President, reflecting on policy concerns in higher education.

In conducting evaluation studies, the CNE uses quantitative methods, such as statistical data, but sees the need also for the use of qualitative methods to fully examine higher education. A typical institutional evaluation involves an internal and an external phase. Following an initial site visit by the CNE team to explain the aims and objectives of the study, the internal phase takes the form of an analysis prepared over a period of two to three months by the institution under scrutiny, using a questionnaire-based survey to assess its strengths and weaknesses. The latter external phase uses a peer review process, with the choice of appropriate peers based on the results of the internal study. The CNE then produces a synthesis based on the results of these two phases. Once a draft of this report has been approved by the CNE it is presented to the head of the institution for comment, and a final draft is prepared, which must be approved by a plenary session of the CNE.

2.3.2 CNER[18]

The CNER (National Committee for the Evaluation of Research) is an interministerial organisation charged with evaluating the implementation and the results of the French national R&D policy (i.e. all the institutions and R&D activities receiving state support from the civilian budget). The CNER can also, in response to Ministers' requests, consider issues relating to military R&D (although it has yet to do so), as well as company R&D.

The CNER was set up in 1990 in a similar fashion to the CNE, independent from the Ministry of Research, reporting directly to the President. Although the two bodies have different

[17] www.cne-evaluation.fr/

[18] Interviews conducted with Professor Linda Hantrais, Loughborough University, and member of CNER and Patrice Van Lerberghe Secretary General of CNER.

objectives and cover different areas of activity, they do work together on some aspects of evaluation.

The CNER made a significant shift in direction in 1996, going beyond the formulation and publication of recommendations concerning the institution evaluated, *"to progressively identify the strengths and weaknesses of the whole national R&D system"* (Laredo *op cit*).

The CNER receives its budget from the Ministry of Research, although this does not in practice compromise its independence. It is answerable to, but not controlled by, 22 government departments. The CNER provides an annual report to the President. Members of the CNER are appointed for a term of six years, although appointments are staggered, with a three-year overlap between groups of members. The CRST plays a role in appointing the members of the CNER, and the OPECRT have a strong interest in their activities.

The choice of evaluation studies conducted by the CNER depends on two factors. First, there is the identification of particular themes, aligned with government research priorities, by the Ministry of Research and other ministries. Second, there are issues identified as important by the CNER, to provide a better understanding of Government policy. In this sense, the CNER's evaluations have, by definition, an objective of strategic assessment.

At the international level, the CNER has recently evaluated the relationship between the French R&D system and the RTD activities of the EU. It examined the French situation with regard to scientific and technological cooperation with developing countries. The international dimension of its work is likely to be extended in the future.

The mission of the CNER is defined by law (1989 decree), which provides it with access to research organisations and permits it to hold hearings. In policy terms, there has been a significant change in the CNER's policy. Since July 1999, the CNER has had two foreign members (one from Germany and one from the UK, who bring to bear experience from the Research Assessment Exercise operated by the Higher Education Funding Council for England, and the Economic and Social Research Council), giving it a broader, more diversified perspective. It now includes higher numbers of active researchers and has also responded to criticism of being too Paris-oriented in its membership.

In terms of assessment of its evaluation procedures, the CNER itself adjusts its procedures in accordance with what it is evaluating. The CNER is obliged to review how its recommendations have been implemented in subsequent years, which can lead to a revision of its procedures. From the external perspective, it must be noted that the recommendations of the CNER carry no obligations for the evaluated organisations. In this sense, an external measure of the credibility of the CNER's work is the extent to which its recommendations are taken on board. Also, the CNER has recently been reviewed by the Parliamentary Office for the Evaluation of Scientific and Technical Choices, a body composed of members of both houses of the French parliament.

There are essentially three sets of units of assessment considered by the CNER:

1) public research institutes and universities (in cooperation with the CNE);
2) inter-organisational and inter-ministerial national R&D programmes ;
3) particular institutions or management and financial procedures in public sector research.

Evaluations are actually directed and conducted by working groups composed of members of the CNER, who determine the methodological approach to be applied and the external experts used.

Research institutes conduct their own internal evaluations (of personnel, etc.), employing their own specialised methods. These are currently the focus of a comprehensive CNER study, which is considering their operation, coherence, transparency and effectiveness. The results of this study are expected in 2001.

Because of the diversity of units of assessment, the CNER has always refused to employ a standard set of criteria, preferring to tailor its approach to each particular set of circumstances. In general terms the CNER could be said to use a multi-criteria approach. The perceived strengths of this approach are its flexibility and adaptability, while its weaknesses centre on the difficulties in setting boundaries to studies.

In terms of the evaluation of impacts, the approach of the CNER is based on the results of R&D. It also uses retrospective analyses, with a time horizon of 5 to 10 years (particularly for academic research). The CNER's recommendations take more of a short and medium term perspective, particularly with reference to where government money should be invested.

Currently the CNER is examining the role played by certain quantitative methods (including bibliometrics) in the evaluation of academic/fundamental research. This interest has been driven by the gradual movement towards global methods of evaluation, where other countries face similar problems in the management of their scientific and technical potential, and the globalisation of R&D activity, driven by international markets for innovation and new information and communication technologies.

2.3.3 CNRS[19]

The CNRS is currently undergoing significant organisational reform, which includes its evaluation section. However, it is possible to comment on some of the key aspects of its evaluation activities.

The National Committee of Scientific Research, which was founded at the same time as CNRS, in 1945, is the advisory body for basic research evaluation. This committee evaluates the 12,000 researchers and the 1,200 CNRS laboratories, through a peer review organisation. The entire field of scientific knowledge carried out by CNRS, has been broken down into sub domains corresponding to "sections" of the National Committee. This division is regularly revised to take into account the developments of science. Each of the current 40 sections has 21 members - 14 elected and 7 appointed by the Minister responsible for research - making a total of 840 members. The voting base is extremely large, consisting of the entire French public research community, as well as those researchers working in private companies in close collaboration with CNRS. The sections meet twice a year for 3 or 4 days in the spring and autumn. The autumn session is devoted to the evaluation of the laboratories and their researchers, the spring session to promotion and recruitment of researchers. The sections also analyse the state of science in their domain and its perspectives in a national and international context. The national committee is composed of sections, which are responsible for the

[19] www.cnrs.fr

evaluation of laboratories and researchers, focusing on their scientific activity. Researchers are evaluated (performance/career appraisal) every two years.

2.3.4 INRA[20]

The Institut National de la Recherche Agronomique was set up in 1946 and became a national public scientific and technological establishment in 1984, under the joint authority of the Ministries of Research and Agriculture. Today INRA has a threefold mission to:

- guarantee consumers high-quality food
- ensure that agricultural and agro-food companies are competitive
- contribute to integrated land development and sustainable management of natural resources.

As with other research organisations INRA was established by decree. In recent years it has undergone a certain amount of evolution. As with other research organisations, there has been convergence with the university sector, with over half of INRA's research units now joint or "mixtes" units.

INRA's activities strongly reflect society's needs, with current interests including genetically modified organisms, BSE, and water quality. These themes also reflect European level concerns, with one quarter of INRA's annual budget (280 million Francs in 1998) coming directly from the European Union. As INRA aims to respond to current needs, there are significant changes in its research objectives over time. Much of INRA's research (about 80%) is applied, with about 20% more fundamental in nature. All of the research funded by INRA is performed intra-murally, although there is collaboration with technical institutes funded by taxes.

INRA conducts evaluation and related activities at three main levels:

1. At the level of the Director General, who, together with the board of 6 Scientific Directors, consider activities in more strategic terms, determining the balance between scientific quality and credibility and policy priorities;
2. At the research department level, of which there are 17 disciplinary oriented departments, headed by senior scientists, matching scientific capabilities and interests with strategic requirements; and
3. At the individual laboratory level, of which there are around 400, with 8,000 staff directly paid by INRA.

Across these levels, there are various cycles of activity. Each year there are internal hearings by the board to assess each department's present and proposed activities. The results of these hearings are publicised within INRA. Each department evaluates its constituent laboratories and their activities every four years, while within each laboratory, staff are evaluated every two years, within a more detailed evaluation every four years.

Projects are evaluated in a similar way, through the laboratories in which they are conducted, on the basis of the results achieved, in terms of the achievement of three main objectives:

---

[20] Interview conducted with Monsieur Olivier PHILIPE Secrétariat Général de l'évaluation, INRA. INRA (2000) INRA, Research and Innovation, Food, Environment, Agriculture and Society.

- Scientific Quality;
- How the laboratory takes account of strategic objectives in research projects; and
- The general organisation in terms of the balance between socio-economic relevance and scientific excellence.

One acknowledged weakness of this approach is the difficulty to make a compromise in the dynamic of the research between what is 'exciting' for scientists and more general socio-economic needs. It is a question of managing a tension in public funding.

In terms of the examination of outputs and impacts, INRA's funding of research is seen to have a life cycle of 2-3 years, in terms of obtaining a good estimate of scientific impact in terms of publication. Longer term, more applied impacts are viewed over a much longer time horizon (10 years or so). Of course, over this period, attribution can become more difficult.

Individuals are evaluated in terms of three sets of criteria:

- the scientific results achieved
- their more general coordination activities
- their technology transfer-oriented activities.

Clearly the balance between these three aspects vary according to individuals, age and career pattern.

The way in which social controls contribute to shaping the research agenda is seen, by INRA, as one of the strengths of its evaluation approaches. However, the processes are also seen as cumbersome, particularly in terms of the committee structures.

In 1999 INRA organised "Chantier Evaluation," a serious internal reflection of the nature and role of evaluation in INRA[21]. This followed a consultation on the overall organisational reform of INRA. Following the establishment of and subsequent report by a working group of 12 researchers and engineers, a larger group of INRA unit heads, department chiefs etc., debated the issues, both electronically and in person. A final synthesis[22] was produced integrating all the conclusions and reflections. Its recommendations centre on a more systematic approach to evaluation at all levels, with the creation of a Secretariat General for Evaluation in order: to provide methodological support for evaluation activities at all levels, to provide regular accounts of evaluations and their outcomes, and to develop methods and procedures. This should be supported by the information system.

2.3.5 INSERM[23]

Created in 1964, the French National Institute for Health and Medical Research is a public scientific and technological organization. Overseen jointly by the French Ministries of Research and Health, INSERM's remit is to promote health for all. Its current mission and structure derive from a 1983 decree. Its units combine INSERM researchers with university and hospital researchers. In the past three or four years there have been closer and closer links

[21] INRA (1999) *Chantier "Evaluation": Les Conclusions du Chantier.*
[22] Chantier Evaluation
[23] Interview Conducted with Claude Graillot-Grak, Head of the Scientific Evaluation Service, INSERM. INSERM (2000) *Inserm 2001: De la biologie a la sante publique*, Press Dossier. INSERM (2000) *Rapport d'Activite 1999.*

with university researchers. This has made evaluation more onerous, but does facilitate the evaluation of university hospitals, which were not formally evaluated before. INSERM directly pays the salaries of its own researchers, but not those of hospital researchers. Its research units are 50% funded by other organisations, including the EU, charities and industry. Research is divided between fundamental and applied. Its activities at the European level have increased to the extent that its evaluation documentation is now bilingual.

While the majority of research (90%) is conducted within INSERM's own organisations (including university hospitals), some strategic research contracts are let to outside teams where it is felt that this will encourage the emergence of new areas (such as in identified gaps or areas of weakness e.g. psychiatry and nutrition).

Annually, along with INSERM's budget, an outline of desired research objectives is defined by the Minister in line with policy aims. Although the Director General of INSERM is selected by the government, it does, however, have a significant degree of autonomy. For example, it can decide how it meets its broad objectives and allocates its budget. It issues its own calls for tender (largely within its community of researchers), which are usually on a thematic basis. It also launches calls for tender in association with other organisations (e.g. AFM). Usually, these are funded on a 50-50 basis.

2.3.5.1 The evaluation of research units at INSERM

INSERM has a dedicated Department of scientific evaluation, which was created two years ago. Its creation reflects a recent accent on evaluation, and the department was reorganised about a year ago, with the arrival of a new Director General. One of the key issues for INSERM is the identification of gaps in research activities, in line with the perception of the Director General, who reports to the parent ministries. However, there is also the familiar tension between scientific excellence, which was the principal driver 4 or 5 years ago, and strategic aims, i.e. the fulfilment of INSERM's stated mission.

The examination of proposals is a two-step process which involves the scientific council of INSERM (30 members from within INSERM, renewed every four years) and, as in INRA, a set of specialised scientific commissions (CSS). There are currently 8 such CSS for some 2,000 INSERM researchers and 260 research units.

Evaluation is conducted at the levels of research activities, research units and researchers. The principal methodology is peer review although there is significant interest in more quantitative assessment techniques, such as bibliometric-based approaches. The peer review process involves site visits as well as the submission of research proposals (in the case of units) to anonymous referees, usually from outside France. Evaluation consumes a significant amount of financial and human resources.

The annul cycle of evaluation includes the spring "concours" for the recruitment of researchers and the evaluation of bids to create new research units and possible promotions for existing staff in the autumn. On the basis of these, the final decisions rest with the Director General.

Individual projects are evaluated by committees of referees, set up on an *ad hoc* basis to judge their scientific quality.

Research units are evaluated every four years, again by *ad hoc* committees, depending on the research theme. Their lifetime can be extended, flowing a favourable evaluation to a maximum of twelve years, although they may continue for longer under a new director or with a new structure.

In terms of quantitative indicators, publications are used, with the impact factors of journals used as one determinant of quality. Although not currently operational, an automated system is being implemented to look at the workflow.

The system for the assessment of responses to calls for tender is also being automated to make it quicker, and it expected to become operational next year. There is no longer-term reflection on the results of INSERM research, assessment criteria are largely scientific rather than applied.

2.3.6 Others

The Centre for Atomic Energy (CEA) has well established evaluation practices. Its scientific activities are evaluated by some 17 scientific councils. In 1997, 9 supplementary scientific and technical councils were set up to evaluate its more applied research. These councils comprise more than 300 experts from a wide range of disciplines. Some 97% of these are external and 32% are from outside France. Under this system all of the CEA's laboratories are evaluated every 2 to 4 years. Each council produces a set of recommendations, and the CEA produces an annual synthesis of these, together with a summary of the evaluation methodology and the strategic orientation of the organisation. This year, a more general evaluation of the whole of the CEA has been conducted for the first time, through a "visiting committee" of international experts.

The Institute for Development Research (IRD) has recently reinforced its evaluation procedures to enhance the scientific quality of its research on the international level, in response to the priorities of its parent ministries. It has proposed the use of an evaluation grid by its expert evaluators. This is composed of four groups of evaluation criteria. These relate to the scientific quality of research projects, their relation to the state of the art in their field, the quality of the formulation of proposals, and their feasibility and credibility.

2.4 Overall conclusions

Although this study has not been fully comprehensive, it has been able to examine the key actors in the French science and technology system and particularly those in the field of evaluation. It can also be inferred that they provide a clear insight into the current trends and issues in evaluation in France. These can be summarised as follows.

Over the past decade or so evaluation has become more important and has become increasingly institutionalised in France. At the national level, the structure and operation of two key committees, CNE and CNER, clearly reflect this. In addition, the major research organisations all take the process of evaluation seriously and have dedicated units and procedures in place. In most cases, these procedures embrace the full range of activities, from the determination of research themes in line with national policy, through their articulation and impact through research departments, units, teams and projects, to the recruitment and appraisal of individual researchers. These operate systematically, according to, in a number of cases, pluriannual cycles.

While national policy provides a set of broad objectives and context for activity, the organisations examined enjoy significant degrees of independence, despite being in receipt of government funds. In the case of the CNER and CNE, their interministerial remit and role requires such autonomy. A number of research organisations also operate under the auspices of more than one ministry, and are relatively free to dispose of their research budget in line with their own internally defined strategic objectives, which are, of course, coherent with national policy objectives.

Another recent key issue is the convergence between university sector and research institutes. This has led to the formation of joint teams, thereby avoiding duplication, facilitating communication and enabling more effective evaluation.

The feedback to the policy process is an important feature of evaluation, at both national and organisation levels. The CNE and CNER have both produced recommendations directly to the government as well as to the organisations under scrutiny.

There is need to find a balance between scientific quality and excellence and the match with socio-economic needs. Although this depends to an extent on the mission of the organisation, the policy relevance of research is an increasingly important assessment criterion.

Institutional evaluation is largely peer based. The value of quantitative approaches is recognised, and in cases peer review is being supplemented by the use of (largely bibliometric) indicators, sometimes as a matter of course. Their use by the Ministry of Research is clearly an influence, with the caveat that bibliometric assessment is still largely biased towards English language publications.

While evaluation is becoming much more systematised, there are still (necessary) informal and *ad hoc* dimensions giving it the flexibility it requires.

The most commonly cited weakness of the evaluation systems currently implemented in French organisations seems to be their heaviness, both in terms of the associated bureaucratic procedures and the work entailed by the numerous committees, peer reviewers and so forth.

There is an apparent trend towards the broadening of evaluation, in terms of the choice of evaluators, both referees and committee members, in terms of discipline and nationality and even regionality, shifting the centre of gravity away from Paris.

Evaluation itself is currently a well debated topic in France. Its role in the research and policy processes and well as the variety of methods used by different organisations in different circumstances are the subject of reports and studies (for example, the INRA 'Chantier" and the comprehensive study currently being conducted by CNER). The fact that in discussions with representatives of organisations, a familiar set of names frequently recurred, suggest that there is the *de facto* development of a network of evaluation experts.

2.5 Bibliography and sources

Chastgnaret, G. and Goujon, M. (1998) Propositions pour simplifier, deconcentrer et mutualiser les functions de gestion de la recherché dans les Etablissements Publics Scientifiques et de Technologie report to the Education Minister.

INRA (1999) *Chantier "Evaluation": Les Conclusions du Chantier.*

INRA (2000) INRA, Research and Innovation, Food, Environment, Agriculture and Society.
INSERM (2000) *Inserm 2001: De la biologie a la sante publique*, Press Dossier.
INSERM (2000) *Rapport d'Activite 1999.*
Laredo, P. (1997) Evaluation in France: a Decade of Experience in OECD *Policy Evaluation in Innovation and Technology: Towards Best Practices.*
Ministry of Education Major French Research Institutions.
Mustar, P. and Laredo, P. (forthcoming) Innovation and Research Policy in France (1980-2000) *Research Policy.*

2.5.1 Other sources

www.academie-sciences.fr
www.agriculture.gouv.fr
www.assemblee-nat.fr
www.cne-evaluation.fr/
www.cnrs.fr
www.education.gouv.fr
www.education.gouv.fr/sup/univ.htm
www.environnement.gouv.fr
www.equipement.gouv.fr
www.obs-ost.fr/ost_fr
www.recherche.gouv.fr

3 European Study Report: Germany

Prepared by Paul Cunningham

3.1 Introduction

Following the remit of the terms of reference for the study, this element of the review focused on relevant activities in Germany. This report presents an overview of the German research system (Section 3.2.1). In gaining an understanding of the process of evaluation (which is taken to encompass the entire spectrum of assessment activities from *ex ante* appraisal, programme/project definition, selection, monitoring and *ex post* review) it is essential to gain a clearer picture of the context within which it takes place. As will be illustrated, this system is highly complex.

Because of this complexity, and also the time and resource limitations of the study (which also encompassed France, Sweden and the UK), it has not been possible to provide a comprehensive account of the measures used by Federal and Länder Ministries, research institutes and other components of the S&T system for measuring and ensuring excellence. Rather, a general overview of the German evaluation scene is presented (Section 3.2.2), followed by a range of specific examples drawn from a variety of research actors and fields (Section 3.3). Whilst these examples cannot be regarded as typifying a national approach to assessment, it is possible to identify some general trends (Section 3.4). In addition, although the Terms of Reference and the SAGE and BEST documents provide a clear indication of the scope of the study, Government S&T in the German context is less easily defined. For these reasons, the study examined a broader variety of institutions in the expectation that valuable lessons might equally be learned outside of the narrower scope originally defined.

3.2 Main Findings

3.2.1 Overview of system

The major feature of the German research system is that it is complex and interwoven, comprising a large number of actors. These actors may be placed into a number of broad categories or groups[24], thus:

- The Higher Education Institutes (universities, Fachhochschulen, etc.) account for the largest share of research expenditure after industry. The major focus is on basic research and long-term application oriented basic research. The majority of their funding comes from the Federal States (Länder), with additional support from the Deutsche Forschungs Gemeineschaft (DFG or German Research Society) – although the share of industrial research contracts undertaken by universities is increasing. The DFG is charged with the promotion of research and, whilst government-funded, is largely autonomous.
- Strategic research - that is long-term oriented research of a speculative or high-risk nature, and research which involves high costs (due to the required equipment and facilities) - is performed largely by the research centres of the Helmholtz Society.

[24] after Kuhlmann, S. (2000).

These "national research centres" have greatly broadened the scope of their research and now cover fields such as high-energy physics, space technology, medicine, biotechnology, applied mathematics and software development, and environmental technology. Around 90% of their funding comes from the BMBF, the remainder from the Federal States.

- The institutes that are most closely linked with the interests of the Federal and Länder Governments are the "government research institutes" which undertake so-called "departmental research". However, it should be noted that, despite being quite closely linked with the interests of Federal Ministries that sponsor them, all exhibit a high degree of autonomy and in several cases have developed strong scientific-academic agenda which may diverge from those of their associated Ministries.
- Research with a strong basic orientation (largely in the natural sciences and humanities) is performed by the Max Planck Society (MPG) institutes. To a large extent the MPG institutes occupy a niche which is not filled by research in universities, either because of its disciplinary nature or because of the resources needed. These institutes receive all their funding from the Federal Government and the Länder.
- The so-called Blue List Institutes (or institutes of the "Science Association G.W. Leibnitz") are jointly supported by the Federal Government and the Länder. These perform medium-term pre-competitive research in the humanities, economics and social sciences, life sciences, mathematics and physical sciences, and environmental science. There are around 80 institutes of which 16 perform service functions for research (information centres, specialist libraries, research museums).
- The Fraunhofer Society (FhG) institutes concentrate on the transfer of technology through long-term application-oriented research and applied research. The FhG does this principally via short-term contract research for industry and government customers. Thus, it receives a small amount of funding from the state as a contract research organisation. The institutes of the FhG are often viewed as an interface between science and industry and the research organisation has undergone a high level of growth over the last three decades.
- The largest share of R&D in Germany is performed by industry – largely in the area of applied research and experimental development. Relatively little long-term application-oriented research is performed and this is mostly limited to the chemical and electrical engineering sectors. Alongside industrial firms themselves, applied research and experimental development is also performed, in pursuit of industry sector requirements, by the research institutions of the Confederation of Industrial Research Associations (AiF). Some of this research, which is geared specifically to SMEs, is publicly funded.

Thus, in seeking to answer research questions whether in pursuit of policy advice, policy formulation, problem-solving, or long-term S&T requirements, the German Government (and the sixteen Länder governments) have recourse to a diverse range of research performers. Moreover, research contracts (the sponsoring of research for the Federal government adheres closely to the "customer-contractor" principle) may be placed with consortia drawn from more than one of the available sets of actors. The fact that individual research institutes, such as the national research centres and the Government research institutions, may also out-source research projects further complicates the picture. Lastly, an additional level of complexity is conferred by the fact that there is no strong, central policy body responsible for

science, research and technology[25]. Although the BMBF has an oversight role, the Länder are essentially responsible for science and academic research. These run almost all the Higher Educations Institutes, maintain (to various degrees) non-university research institutions and some have launched their own technology policy programmes. As Kuhlmann (1995) notes, whilst this leads to some redundancies in the capacities of the research and innovation system, it guarantees a decentralised autonomous structure of research capacities.

Within this environment, the activities of the Wissenschaftsrat (WR - Science Council) are of significant importance. The WR is an independent science policy advisory council which provides informed advice to the Federal and Länder Governments. It is tasked with the preparation of reports and recommendations on the structural development of universities, Fachhochschulen and research institutes. To assist in this task, the WR conducts evaluations, generally at the institutional or "system" level.

Because of these complex interactions and interdependencies, coupled with the high degree of institutional autonomy and independence, it has proved difficult to identify clearly defined performers of "research in support of Government policy". The concept of a 'Ministerial line-department' is relatively unfamiliar in the German context. The closest approximation to the 'Government laboratory' model is found in the National Research Centres and the government research institutes. However, the research activities of these institutes are not limited to specific Government requirements. Neither is the performance of Government-oriented research limited to these institutes alone. The complexity of policy questions and the research advice which is needed to address them frequently requires the input of a range of research capabilities which may exceed that of dedicated departmental research facilities. Indeed, given the multi-disciplinary, ever-expanding and rapidly developing nature of today's scientific issues, it is almost impossible for a single institute to maintain a comprehensive expertise even in a number of closely related fields (such as crop protection). There is thus greater reliance on an interconnected network of research performers which is able to respond to these broad demands. Moreover, a narrow reliance on a set of 'dedicated, mission oriented' research institutes would ignore the available strengths of the multi-faceted national research system. These tendencies are thus in accordance with the conclusions of Gibbons *et al*[26] in their assessment of the shift of research and technology knowledge production from "Mode 1" to "Mode 2". That is to say, in seeking answers to its policy requirements, the Federal and Länder Governments access the research system using an approach which is application oriented, transdisciplinary, heterogeneous and heterarchical.

3.2.2 Overview of evaluation

Given the complexity of the scientific system, it is unsurprising that the German evaluation system has developed an equally diverse set of approaches to address the problem of ensuring research excellence (in its broadest sense).

Briefly, there are several levels at which evaluation may be applied[27]:

1. (Institutional) structural level. For example:

[25] Kuhlmann, S. (1995)

[26] Gibbons, M., Limoges, C., Nowotny, H., Schwartzmann, S., Scott, P. and Trow, M (1994), "The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies", Sage, London.

[27] See Uhlhorn, C. (2000)

- The Wissenschaftsrat examines and evaluates systemic questions, such as the level of cooperation between universities and the national laboratories (Helmholtz Centres).
- Specially convened peer review committees undertake "system evaluations" (such as of the MPG, the DFG and the FhG). These may involve the use of international commissions. The Wissenschaftsrat has also examined the Helmholtz Centres and all important institutes in the field of environmental research.
- Both the Federal Government (under the legal framework act for universities - Hochschulerahmengesetz) and the Länder (under the Landeshochschulegesetze) have induced structural changes at universities. These processes have been guided by expert advice and extensive discussions between various interest groups.

2. Individual R&D institutes (either regularly or on request). For example:
   - The Wissenschaftsrat is in the process (since 1994) of evaluating the WGL (Blue List) Institutes. Its recommendations cover closure, restructuring, down-sizing, or continuation of financial support.
   - In the 1990s, the MPG assessed and reformulated its evaluation procedures. It retained peer review – *ex ante* assessment of individuals' scientific potential – as its primary evaluation mechanism.
   - In keeping with its orientation towards application-targeted research, the FhG largely employs indicator-based approaches (e.g. contract money raised, cooperations, etc.)
   - The Helmholtz Centres operate their own individual evaluation practices, some of which were formulated in the 1970s and subsequently modified. The BMBF has also requested a number of peer reviews of various Centres and specific external policy advice has been provided concerning large investment decisions.

3. Broad objectives and research goals.
   - An emphasis on clear programme definition, through a broad range of *ex ante* assessments, appears to be a major feature of German research. These may be conducted by individual research organisations or may be requested (and, possibly, funded) by the BMBF. A large number of *ex post* evaluations are also conducted.

4. Assessment of support and funding instruments (especially measures dealing with SME support, Start-up support, and cooperative research projects).
   - The "call for tender" approach used by Ministries and individual research institutes represents a type of system evaluation for programmes.

Kuhlmann (2000) characterises German evaluation practice in the area of judging the performance of research and research institutions as both strong and fragmented. Its strength emerges from the high degree of self-organisation which encourages consensus and commitment amongst researchers. However, because few efforts have been made to coordinate institutionally oriented evaluation, it is unsystematic and fragmented. He identifies three levels or "shells" for evaluation:

- At the "core" level, the research performance of individuals and research groups is measured, using peer review and other processes (such as bibliometric studies). Peer review is widely used in the German research system (notably by the DFG) and particularly in the *ex ante* evaluation of projects in basic and long-term application-oriented research. In the case of the DFG, applications for grants from individuals or groups of researchers are assessed by a "college" of peers. These are elected by the entire scientific community and serve for a period of four years. Assessments are based purely on the scientific quality of the application.

- The second shell consists of evaluation studies or impact analyses of research and innovation policy programmes. These analyses are generally initiated by the "policy administrative system" (e.g. Government) and have the objective of verifying the attainment of politically set, scientific, technological, economic or societal goals. Such studies may also incorporate efficiency reviews. The Federal Ministry for Research (then the BMFT) has been a major user of these types of study, commissioning in excess of fifty large-scale evaluation studies. Such studies are generally conducted by independent research institutes using a broad spectrum of evaluation concepts, methods and instruments. Collective experience in these studies has led to the development of a German-speaking evaluation community.

- The third shell consists of the evaluation of the performance of entire research institutions. Numerous studies of individual institutes or system evaluations have been undertaken or are still ongoing, including those mentioned under level 2 of the preceding classification.

As might be expected given the complexity of the S&T system, the practical use and functions of evaluation procedures are wide ranging. They may also be performed for a variety of purposes, depending on their subjects and customers thus[28]:

- for legitimising the distribution of public money and demonstrating effective use of funding;
- for targeting (selection of priorities) and controlling ("fine tuning"), to aid in improving the management of S&T policy programmes;
- to improve transparency with regard to conditions for participation and expenditure;
- to enhance the information basis for policy development and formulation.

Kuhlmann (ibid.) also sees evaluation acting as a moderator "between the diverging and competing interests of the various players in the S&T system".

3.3 Specific cases

Due to the complexity of the German S&T system outlined above, and due to time and resource limitations, it was not to provide a comprehensive, in-depth view of the entire national system for the assessment of research. Indeed, there is no such thing as a national system, or even, for that matter, a standardised ministerial system by which policies are formulated, programmes defined and implemented, projects assessed, monitored and evaluated. The autonomy and independence of the research institutes coupled with their

[28] Kuhlmann (forthcoming)

varying modes of action and fields of activity, militates against the development of such a unified system.

A number of specific case studies of ministries and research institutes are presented below.

3.3.1 BMWi - Federal Ministry for Economics and Technology and BAM (Federal Institute for Materials Research and Testing)

The **overall role** of the Federal Ministry of Economics and Technology is "to lay the foundations for economic prosperity in Germany spread broadly throughout the population" (http://www.bmwi.de)[29]. There are three main areas for research programmes – Energy, Aerospace and Communication Technologies. Other areas include Biotechnology and Genetic Engineering. A large proportion of BMWi's budget goes on helping SMEs keep up with developments in innovation. A major criterion concerns the participation by research organisations and firms of the former DDR.

BMWi oversees activities of the PTB (Federal Institute of Physics and Metrology), the BAM (Federal Institute for Materials Research and Testing) and the BGR (Federal Institute for Geosciences and Natural Resources). These have their own research facilities. The Ministry has no internal capacity for its own research. Some research is contracted out, but most research funding goes to SMEs and research institutes.

Priorities for research are generally decided on an annual basis but are heavily influenced by the preceding year's priorities. Decisions are made internally - there is no external advisory body. However, decision-making is a hierarchical process, made at the branch level, then the ministerial level and with final sanction from the Finance Ministry. Whilst all departments have an idea of their particular priorities, the final decision is arrived at through an integrated process of negotiation. Priorities for spending are set at the permanent secretary level (based on political judgement). Selection at the next level down will be made, for example, by the Energy Department.

BMWi uses an approach similar to that of the UK ROAME-F system for **programme definition**. Programmes start with an *ex ante* evaluation (using experts). The ministry is now embarking on a more ambitious "systemic evaluation" procedure (see Belitz, H. and Lorenzen, H-P.)

Programme evaluation is conducted by external contractors who compete in an open tendering process. Intermediate workshops are used to ensure that the requirements of the evaluation study are being met.

BAM (the Federal Institute for Materials Research and Testing) is a technical and scientific Superior Federal Institution under the authority of BMWi. It has a staff of about 1,700, including over 700 scientists and engineers. It comprises 10 Departments divided into 34 Divisions with 127 project groups, laboratories and sections. Annual budget is about 190 Million DM (Federal Funds) with a further 32 Million DM (Third Party Funds, Fees).

[29] An overview of the policies of the BMWi may be viewed in "The Courage to Innovate, Technology Policy – Paths to growth and employment".

The task of BAM is to promote the development of the German economy in regard to the fields of Materials, Chemistry, Environment, and Safety. In particular, it is responsible for:

- statutory functions for technical safety in the public domain, especially relating to dangerous goods and substances
- collaboration in developing legal regulations (e.g. safety standards and threshold values)
- consulting on safety aspects of materials technology and chemistry for the Federal Government and industry
- development and supply of reference materials and reference methods, especially for chemical analysis and materials testing
- assistance in developing standards and technical rules for the evaluation of substances, materials, structures and processes with reference to damage prevention, life time prediction, and protection of economic values.

Towards these goals, BAM undertakes R&D, testing, analysis, approvals; and consultation and information dissemination.

The tasks of BAM require interdisciplinary cooperation. BAM collaborates closely with technological institutions in Germany and abroad, especially with national institutes. It gives advice to Federal Ministries, economic associations, industrial enterprises and consumer organizations. It provides expertise to administrative authorities and law courts. In the area of measurement, standardization, testing and quality assurance BAM is the competent national authority for testing techniques. BAM also cooperates with numerous technical, legislative and standardization bodies in order to develop technical rules and safety regulations and represents Germany at both the national and international level.

There is no direct external control over BAM's activities – it is largely autonomous provided it adheres to its remit. However, the institute has been evaluated externally by an independent company (at the behest of the BMWi) about 5-6 years ago. A Kuratorium, composed of external scientists, industrialists, BMWi officials, and staff of the Ministry of Economic Affairs, assesses BAM's work and activities and holds an annual board meeting. However, BAM has oversight of the membership of the Kuratorium, thus there are no "unfavourable" members.

BAM undertakes Programmes and projects. Projects usually have a limited lifetime of about three years, although they may be prolonged in certain circumstances. Projects are approved by the BAM Presidential Board (an internal process). Programmes have to fall under so-called "Main topics", i.e. relate to the safety and reliability of materials and or chemistry under the general headings of: analytical chemistry, materials technology, environmental technology (a detailed methodology for the organisation of project applications is provided in the "Projektsorganisation Manual – Teil 1").

Projects are selected according to criteria relating to:

- effectiveness
- efficiency
- contribution to safety and reliability
- impact on regulations and standards

- subsidiarity[30]

A relatively sophisticated scoring system is used to assess applications on a project by project basis (see "Projektsorganisation Manual – Teil 2"). Each of the criteria may be allocated up to three points to give a maximum score of 15 points. A further criterion which may be applied is the relevance to the needs of SMEs. Applicability to the needs of customers forms a key issue in programmes of this nature. A translated example of the project scoring guidance and results sheet is provided below as an Annex to this country report.

There is a control system for project management: Project Heads submit a report every second year, plus a final report. These reports should mention all successes, including the amount of external funding attracted, numbers of papers produced, standards and reference materials resulting, new laws, etc. Projects are assessed by the Presidential Board after a special Working Group has studied them and prepared a report on each one. Roughly 300 projects are underway at any one time. The Working Group is also responsible for reviewing all project proposals before they are passed to the Presidential Board. Monitoring of projects is conducted via the two-yearly reports

Completed projects are evaluated by means of the final report, produced 4 months after the end of the project, which is reviewed by the Presidential Board. Final Reports are expected to take account of outputs (such as patents) but there is no method for assessing outputs subsequent to the final report. One problem with the present methodology is that it relies very much on qualitative judgements rather than on quantitative indicators or fully objective measures. Quantitative outputs are recorded via the Project report sheet. These include: laws, regulations, EU guidelines (major or lesser input); orientation to legal regulations (leading or lesser input); norms or technical regulations; validation of processes; reference materials certification; evaluation of laboratory competence; external advisory services; written advice (extensive/normal, amount of time); verbal advice (extensive/normal, amount of time); patent registrations; patent issues; licenses; lectures; publications (in periodicals, monographs, published research papers); awards and rewards; PhDs; MScs; Reports (with monetary value if possible); testimonials; others.

The same evaluation process is applied to all projects regardless of the customer for the research (all BAM work is performed for customers). The BMWi forms a major customer (for new legislation, standards, new testing methods, compliance with EU laws). Big projects may involve large numbers of partners but the same appraisal methods are applied. Slightly different (academic) criteria are applied to projects which involve PhD students (BAM has around 50 such students at any time).

All the procedures in place have been introduced within the last five years, with some minor changes. For example, greater attention has been focused on monitoring the amount of staff time spent on projects. Thus time keeping/auditing has assumed a greater prominence. This has been introduced to allow better managerial control over projects in the face of declining staff numbers and the need to prioritise tasks and allocate funding to priority projects.

[30] Subsidiarity is a term which is generally understood to refer to the principle that central authorities (for example, the European Commission) should only perform those tasks which authorities further down the hierarchy (for example, EU member countries) are unable to perform themselves. In the Canadian context, the principle might be applied to deciding which tasks should be undertaken at the Federal Government level or at the Provincial Government level.

Technology changes have also influenced the type of work undertaken (e.g. nanotechnology, MEMS, image analysis, advanced materials).

3.3.2 BML – Federal Ministry of Food, Agriculture and Forestry

In addition to the task areas suggested by its name, the BML is responsible for a number of other fields at the federal level. It also aims to make a contribution to the sustainable development of rural areas beyond agriculture in the narrower sense. The Ministry coordinates and represents the interests of the Federal Government and Länder within the EU in matters relating to rural areas, agriculture and forestry, viticulture, horticulture as well as fisheries or when problems concerning plant and animal health must be addressed. (see http://www.bml.de/englisch).

The primary objectives of the BML's research are to support policy and policy decision-making/formulation. As with other ministries, the BML is autonomous with regard to decision making and formulation of programmes. Some auditing control is exercised by the Federal Accounting Agency. Advice on programme formulation is also provided from commissions comprising academics, consumer organisations, food industry, etc.

There are several institutes under the control of the Ministry, covering a number of fields of research (see www.bml.de and www.dainet.de). All BML institutes are under Federal control – some are under the BML, others may be co-funded with other ministries. The Länder may also be involved in the control and funding. The overall BML research budget is DM500 million. Some BML institutes are concerned with validation, control, regulation, certification (e.g. BBR), etc. Other areas concern risk analysis (of GMOs and pesticides, food safety). Out of the 10 federal research centres within the scope of the BML, there are two centres (BBA and BFAV) which – besides research – follow legally determined duties, such as authorization of plant protection products, release of GMOs and authorization of vaccines.

As a guideline for the scientific fields to be covered by the federal research centres, BML in co-operation with the research centres establishes a framework of scientific topics ("Forschungsrahmenplan"), which remains in force for 4 years in general. Within this framework, the scientists of the research centres are free to working on scientific questions of their choice. Thus, there is no direct dependence from the government. Nevertheless, there is a close contact between BML and the research centres, as the research centres provide extensive scientific advice at the request of the ministry. This contact provides the research centres with a clear picture of the Government's political needs and objectives. The relative freedom of research provides an opportunity to maintain research topics that are not directly necessary for short term policy goals, but which may become relevant in future (e.g. BSE, plant protection, certain aspects of nutrition research).

Although there has not yet been an external evaluation of the research centres by the Wissenschaftsrat, the quality of the work carried out in the research centres is documented, for example by the relatively high amount of external funds (DFG, EU) gained by the institutes and , of course, by the quality of advice given to the ministry. The fact that several research institutes run laboratories which serve as national reference centres (e.g. for certain animal diseases) provides an additional indications of quality.

The institutes are subject to an evaluation when the director changes. On this occasion a scientific commission, including members from the research centres as well as from other institutions (universities etc.), produces a scientific report on the relevant institute.

Each research centre has a scientific council with members from various scientific institutions, administration and practice. The scientific council advises the research centre on structural problems and carries out its scientific development. Besides the scientific output of the research centres the BML provides funds for special projects in universities and for environmental research projects.

There was a change of policy accent following the change in Government in 1998. This led to greater interest in environmental issues. There is a degree of overlap between BML and the ministries of health and environment – for example, the BMU sometimes get BML institutes to do research. The BML institutes can be commissioned to do research (by other ministries, industry, EU, etc.) provided the research adheres to ministerial guidelines.

The research funded by the Ministry is generally not applied research – although some near-market research may be funded (provided it is not concerned with the interests of a single company). Most research funded is strategic/generic.

3.3.3 BMBF - Federal Ministry of Education and Research

The Federal Ministry for Education and Research occupies a central position in the German S&T system. However, as Kuhlmann (2000) notes, its major role is to act as an interface between science, industrial research and general state science and technology policy. It does not direct nor control S&T policy and strategy, but rather acts as a formative moderator, albeit as a partisan of its own institutional interests. All BMBF research, ultimately, is aimed at societal improvement.

Overview

The BMBF is composed of Referats (Departments) which cover different application areas (biotechnology, genomics, microsystems, electronics, health, traffic, environment). Each department has its own criteria for programme definition, derived from their specific remit and research fields. Whilst they have similar general goals, the precise criteria differ. For example, research policy in the field of health will be set by a council of relevant experts, managed by the BMBF. Representatives of the BMG will also have an input. Recently, a research gap has been identified in the area of clinical research and attempts are being made to link basic research to clinical practice. Genome research is also an expanding area and a new Department has been set up in this area.

The units of assessment are hierarchical and are defined within 'framework programmes' (Rahmenprogramme) which cover a 5-10 year life span. Lower level activities are defined as 'key actions'. At the lowest operational level, reviewers/experts are used for project definition. For example, the programme "Production Technologies" is an 8 year programme, defined by various important sub-areas (see below).

A new systemic approach to programme definition is being adopted. Known as FUTUR, it attempts to give more transparency to the process of defining the rationale for projects. The aim is to find a broader approach which involves discussions with a range of external representatives (including users). Open discussion will be used to define the topics which the Federal Government should be tackling. This represents a new rationale for funding priorities. Currently, rationale is decided at the political level (e.g. based on Kyoto protocol objectives). Otherwise the programmes tend to derive from previous programmes. This lacks transparency. FUTUR intends to improve societal impact - in a somewhat similar to the

Foresight process, and will utilise a website for public debate on the internet. FUTUR is still due to start, but little is happening at present.

Once projects have been defined, open calls for tender are published by the Ministry. Research proposals are assessed by a board of reviewers (supplied with the appropriate criteria by the Ministry). Criteria may also be set by monitoring and project management organisations. These set specific assessment criteria such as: likelihood of conversion into practice, questions related to the implementation of results, expertise and track record of the proposers. Most calls for proposals also specify joint industry/academic collaboration. Objectives and milestones are set by the Ministry and by experts for both programmes and projects.

Third party organisations - 'Projekttrage' - are used to manage projects, transfer money, and undertake monitoring. Evaluations are often conducted by a third agency (for example, relevant institutes of the Fraunhofer Gesellschaft). The reason for this is that the work cannot be done by Ministry personnel – there are too few – the system is described as the extended or "long desk of the Ministry". Most Projekttrage are public bodies thus they are still almost civil servants. (5% of project moneys go for management overheads).

Each topic/project lasts 3-5 years. The final evaluation is conducted by third parties or the Projekttrage and external experts. These may be based on final reports and oral presentations. Projects are evaluated at the topic level and also, depending on the programme, at the individual project level. A yearly report is also expected. This is generally checked either in-house or by the Projekttrage. At the mid-term, a longer report is produced. This goes to external reviewers (from industry, academia, other project or topic areas). These may operate individually or as a group. The final report can be an evaluation of the topic area into which it feeds – alternatively it may just be subject to an administrative evaluation. Topic area level evaluations are done by external experts. The results of the final evaluations feed into discussions for the definition of new topics. The process bears some resemblance to the UK's ROAME-F system.

Monitoring of the longer term impact of the research may or may not take place. After the topic level evaluation there are often suggestions for the evaluation of the dissemination activities. Also the evaluators have the knowledge of the research teams' past expertise. There is no systematised approach however.

The German system is characterised by many departments, etc. which develop their own programmes. Close contact is maintained with the research community via the use of experts. Its strengths include the close contact maintained with the scientists and industrial researchers by the project management organisation. The 30 departments of BMBF use around 700 Projekttrage. This represents a very large network. The main weakness of the process is that it is very hard to set priorities across the whole range of technologies and applications areas. This can lead to interdepartmental competition over budgetary resources.

3.3.3.1 Programme example 1: 'key technologies' programmes.

At the programme level, potential areas for research funding result from discussions and consultation with external experts in the field. Specific areas form the subject of detailed investigative studies conducted by external consultants. One of the key criteria in the definition of programmes is the potential for industrial application. In addition, industrial representatives and academics will be approached on an informal *ad hoc* basis to further

investigate the potential of the area in terms of industrial applicability. Programmes are thus built up from a series of discussions. Overall the field should have potential for broad application (beyond the interest of a single firm) but should also have some connection with other fields and show inter-disciplinarity. All research must be pre-competitive. Another criterion is that the research must involve a certain level of risk (the purpose of BMBF funding research in these areas is to reduce this risk to industry).

If industry has an interest in the area, it generally undertakes the research in conjunction with other users (academics) or with other research institutes. Cooperating firms make a 50% financial contribution.

At the project level, fields are defined using the detailed process above. Open invitations to submit proposals are then published. Proposals are subjected to peer review by independent experts (which may include international experts) and projects selected accordingly (see above for selection criteria). From a priority list, projects are selected until the available budget has been allocated.

Project monitoring (essentially management and also administration) is undertaken by external bodies (Projekttrage). These hold meetings and seminars with the researchers, define milestones, etc., and report regularly to BMBF. Financial input is also monitored. In the case of key technologies, most monitoring is undertaken by the Association of German Engineers (VDI) in Dusseldorf.

Industrial programmes are subject to 5-yearly monitoring evaluations conducted by external consultants. On the basis of the submitted report, decisions are then made as to whether to continue the programme or to terminate it. A specific condition of project funding is that all programmes have a limited lifetime of between 5 to 10 years. Programmes within National Research (Helmholtz) Centres may last longer, and are regularly evaluated every 4 to 5 years. Over recent years, a number of these centres have been forced to close or find new areas of research activity. Some have thus become multi-purpose rather than restricted to a single field of study. There are 16 National Research Centres undertaking research in a range of scientific fields (see http://www.helmholtz.de).

One of the major weaknesses with the process (in some technical and scientific areas) concerns the difficulty in finding high-level experts who are independent of the funded projects and programmes.

Completed programmes are not generally evaluated, although some may be - there appears to be no rule or principle governing their selection. However, the monitoring aspect is very strong. Technology Assessment (impact) studies are also used to examine the longer term consequences of the research at programme level. In the case of completed projects, an assessment is made of the prospects and longer-term effects emerging from the research. The usual monitoring agencies (Projekttrage) are used to carry out the final evaluation which is broadly based on an assessment of whether objectives have been met and whether the stated intentions of industry to use the results have been fulfilled. Some monitoring of outcomes may also be carried out subsequent to the end of the project, but there is no systematic selection of projects for this process.

Over recent years the use of the monitoring process has been increased due to favourable experience of its use. As a consequence the Ministry now supports a number of in-house specialist staff.

3.3.3.2 Programme example 2: 'Production Technologies' Programmes

The selection of "Production technology" as a BMBF programme area was made at the political level – but based on input from a range of expert advisory sources. Germany has a very high dependence on "established" industries which rely greatly on production technology, e.g. automotives, electrical goods, machine tools.

In general, BMBF programmes are defined through consultative studies performed by industry and research institutes. An example is set out in the publication 'Produktion 2000+", published by the laboratory for machine Tools, RWTH, Aachen.

Programmes are sub-divided into work areas, each of which is covered by one or projects. Calls for proposals are then formulated and published. Bids generally come from industry and research consortia. The selection process is supervised by neutral external experts. Each proposal is peer reviewed by 4-5 experts and given a priority ranking. Projects are then funded according to budgetary availability.

Specific criteria are set out in the Rahmenskonzeptforschung fur die Produktion Morgens. In some specialised fields, technical criteria are used. All proposals must be collaborative, i.e. involve academic and industrial participants. Many of the FhG institutes undertake projects under the Production technologies programme.

One problem with the peer review process is the difficulty in locating independent peer reviewers – however, as the field covered is so broad, this is rarely a major inconvenience. It depends on which particular area is under review.

Not all projects are monitored, although larger projects are monitored using contracted external experts (Projekttrager). In order to aid the monitoring process, proposals are carefully prepared according to specific guidelines and contain clearly defined objectives, milestones, financial plans, specific technical aspects and anticipated deliverables. Generally the Forschungszentrum Karlsruhe is used for the monitoring of Production Technology projects. Projects may be very complex, involving up to 50 partners although an average is 7-10).

Sometimes a final evaluation is undertaken. This examines the uptake of results – dissemination of the work undertaken is a very important aspect. This is a relatively new aspect (implemented for 2 years now). Contracts require precise statements of what will be delivered, what will be done with the results and when. There have been no results as yet as no projects have yet been completed using this process (usual project lifetime is three years). Another change concerns the ownership of the results. Rather than being publicly owned, the results now belong to the industrial sponsors who are responsible for bringing them to market. Some public usage rights are retained (such as patent returns) but most of the benefit now goes to the companies who fund 50% or more of the project costs. Again the primary objective is to reduce the risks to industry.

3.3.4 BMG – Federal Ministry for Health

The Federal Ministry for Health is responsible for shaping central elements of the social state and consumer health protection. Together with its institutes, it contributes towards health promotion and warding off hazards to health. The Ministry supports research into new care provision structures, addiction control, disease control and for the purposes of regulation and licensing in areas such as medical and nursing care, and human and environmental protection. (See www.bmgesundheit.de). The BMG does not have particularly strong direct links with the Higher Education and other research sectors. These links are moderated through the BMG's links with the BMBF.

Interviewees noted that it is difficult to explain the German system in the area of health research. National research priorities, including the field of health research, are set by the BMBF. In parallel, the BMG has its own sets of priorities and, through negotiation and discussion, exerts an influence on health priorities in the BMBF's research portfolio. As a consequence, some areas of health research are funded jointly by BMBF and BMG, others by the BMG alone. BMG also funds "special" or "departmental" research necessary, for example, in the formulation of regulations and guidelines in the health sector. The departmental research budget is much smaller than that which goes into the BMBF/BMG funded programmes. There have been no major policy or structural changes since the change of government in 1998 – although there is perhaps a greater interest in alternative medicine, patients rights, etc.

BMG largely funds applied research (mostly in support of regulations, laws, etc.). Very little basic research is performed. Six institutes work on behalf of the BMG: Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) (Federal Institute for Drugs and Medical Devices); Robert-Koch-Institut (RKI) (detection, prevention and control of diseases); Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin (BgVV) (Federal Institute for Health Protection of Consumers and Veterinary Medicine); Paul-Ehrlich-Institut, Bundesamt für Sera und Impfstoffe (PEI) (Federal Institute for Sera and Vaccines; Deutsches Institut für medizinische Dokumentation und Information (DIMDI) (German Institute for Medical Documentation and Information; and the Bundeszentrale für gesundheitliche Aufklärung (BZgA) (Federal Centre for Health Education). These are essentially independent and semiautonomous, with little ministerial influence. They carry out a range of research activities, some of which can be basic-oriented. Each has its own separate budget.

Research funded by the BMG is largely performed externally – in universities, Blue List institutes, epidemiological study organisations etc. Universities perform most of the research. A small proportion of the Ministerial budget funds research in the BMG institutes, usually via public invitations to tender. Research is generally performed for the purpose of informing policy and in the establishment of regulations and codes.

Each of the Federal Ministries tends to have its own system for programme definition, budgetary allocation, project monitoring etc. However, some coordination does occur between ministries. For instance, all projects with budgets in excess of 200,000 DM are notified to the other Ministries in order to ensure that duplication is minimised. Only after other ministries have agreed the project can it proceed.

The divisional representatives define programmes. They receive advice from relevant institutes and experts. The Ministry formulates the specific programme, which is then

assessed and examined by external experts. These experts, who are often identified by the Wissenschaftsrat, usually monitor the programme also. While this is not standardised, it is general practice. The number of experts varies according to the size of the programme and the nature of the task. The assessment process is made on the basis of scientific criteria. The ministry is structured into divisions and sub-units. Sub-unit heads are responsible for both programme and project definition. Project proposals are assessed at a yearly ministerial meeting. Each proposed project follows a clearly structured (but not identical) format which sets out the funding, planning arrangements, etc. A specific Ministerial unit assesses these plans. Each is discussed and usually around 50% are selected for funding. Assessment criteria include the perceived need for the project, the track record of the proposed researchers, and a sound rationale. The ministry then sets the project timetable, and specifies its objectives and anticipated outputs.

The assessment of whether or not the Ministry is obtaining the best possible research, or if it is done by the best people is recognised as a problem issue. At the pre-project stage, the tender procedure generates responses and the long-term experience of the Ministry staff is useful in judging the capabilities of the proposers. In addition, potentially successful tenderers undergo a formal interview and negotiation process which allows judgements to be made. At the end of a project, a report may be produced and sent to an institute or an expert for evaluation. These are often assessed on the extent to which the research contributed to the formulation of regulations.

Within the BMG research institutes, an internal research council discusses directions and priority setting. Each institute has a Forschungsbeauftragter (Commissioner) responsible for the coordination of research.

3.3.5 National Research Centres (Helmholtz Society Centres)

In Germany, research in support of Government policy (for example, into BSE) may be conducted by one of three routes:

- It can be carried out in Government research institutes or laboratories with Government or Länder funding;
- It may be funded by the BMBF – such project funding is open to all applicants;
- It may be performed by the Helmholtz Centres (NRCs), following discussions to determine which topics they could support.

Thus, policy relevant research could be performed by the institutes of the Ministry of the Environment or the Ministry of Agriculture. Similarly, research for food safety monitoring might equally be performed in the Blaue Liste institutes.

The NRCs do long-term research (such as on fusion, at Munich) or on long term monitoring (Augsburg), or health research. However, one weakness of the, autonomous, Helmholtz system is that it does not respond quickly to the needs of Government. Thus Government endeavours to direct their research portfolios through the provision of additional funding in specific topic areas. In addition, the BMBF has oversight of the Helmholtz laboratories, a situation which may cause problems. For example, the Health Ministry may wish to see research conducted in a particular area which is outside the BMBF's immediate interests and the scope of the Laboratories. The BMBF sets broad goals for the Helmholtz, (e.g. health, environment), although in general the research portfolios of the NRCs are not strictly relevant to the type of research conducted in support of S&T policy. Some of the NRCs conduct

monitoring activities, for which they receive additional Government funds. However, basically they are semi-commercial research laboratories and the Government forms but one of their customers.

The Helmholtz Centres are currently the subject of a series of systemic reviews by the Wissenschaftsrat.

Each Helmholtz institute sets its own priorities (all are fairly autonomous). However, some cooperation does occur to avoid duplication. (According to the WR evaluation, this process is not as effective as it could be). Research projects are selected based on the decision of the Senate of the Helmholtz Society. This sets topics for the institutes, and attempts to avoid duplication and enhance cooperation between the centres, although some parallel topics may be allowed provided different approaches are followed (this represents a recent change). Helmholtz networks (in areas such as climate, environment research, geosciences, energy, health) hold discussion meetings to decide on project selection. Groups from the centres then decide internally on the specific subjects, although some discussion with ministry officials may occur (generally Ministry officials do not interfere with the decision making process).

All Helmholtz programmes and projects have been monitored (since 1975). This is performed by review boards using external referees. Four years ago the centres began to set up topic review boards (for topics such as Human Genome, soil water, climate) which covered all the Helmholtz institutes. These examined single topic areas but across all the laboratories. This was seen as an improvement.

Internal monitoring of projects differs between centres – some centres may not monitor projects at all. At the project definition phase, specific milestones and deliverables are set, which provides an aid to subsequent monitoring (if used). However, it is expected that this process will undergo some drastic changes in the near future.

The Max Delbruck centre for heart research in Berlin is regarded as a good example. The research projects are precisely defined and closely monitored. Research projects are defined through a bottom-up process. However, although the centre has some good research groups, they are not too dissimilar from university-based or MPG research groups.

Another good example is the Cancer Research Centre. The Centre uses committees of university researchers to referee its various activities. Every four years an external review board examines the performance of the Centre and it also undergoes a complete internal review every two years. While the primary purpose of the review is for monitoring and quality assurance, it also helps to develop an improved awareness and understanding of the research activities of other research groups in the centre.

At the overall Helmholtz level, there have been few evaluations of completed research. Although the Helmholtz Senate boards are supposed to review past activities, the Wissenschaftsrat is unconvinced that the process is that effective at the programme level due to the scarcity of defined objectives. The Wissenschaftrat's view is that it would be better if the Helmholtz labs reported to an independent body rather than those directly affected by their advice and findings.

3.4 Overall conclusions

3.4.1 Programme definition

Within the Federal Ministries, programme definition generally emerges following a process which combines top-down broad political objectives, external (international) policy demands, and perceived societal needs, within the parameters of the Ministry's remit, with a bottom-up set of advice from industry, academia and other relevant scientists. Ministry officials have the ultimate responsibility for the definition of research programmes, but the process is very much consensus oriented – and may become more so within the framework of the FUTUR initiative. Within the various research performing institutes, programmes are generally defined by a central council – again usually advised by internal, and possibly, external experts.

3.4.2 Project definition

As in the case of programme definition, the formulation of specific projects appears to be a largely bottom-up driven process, based on a, sometimes extensive, negotiation procedure which utilises advice from scientists, industry representatives and other external experts or interested parties. Within Ministries, those Ministry officials directly concerned with the specific area of research in question may define projects, or projects may be more loosely defined and opened to external tenders. In the research performing institutes, project definition may be left to specific managers, or may be defined at a higher organisational level.

A notable feature of the project definition process is the emphasis placed on planning. This appears to take place generally either at the project definition phase or following the proposal acceptance stage. The planning process entails the definition of clearly defined goals and objectives, identifiable project milestones, specified deliverables, etc. Whilst this clearly forms a useful attribute of project management, it may be prompted largely by the need to "hand over" project management and monitoring to a third party (see below).

3.4.3 Proposal/project selection

All selection processes are strongly based on peer review, although the composition of panels or committees may be extended to include industrialists, users and other government officials. The specific nature of the process generally depends on whether a project has been internally defined (by Ministry officials, or by institutional research managers) or is open to external tender. In the former case, a committee review process is generally used to allocate projects according to budgetary availability. In certain cases, a highly refined and structured review process is used, which may employ scoring systems based on specified criteria, highly relevant to the nature of the project.

In the case of projects that are submitted for open tendering, project applicants are subject to an assessment process which employs sets of appropriate selection. The track record of the applicants forms a major factor in this selection process. A two-stage selection process may also be used wherein written proposals undergo a preliminary selection round, followed by a more detailed interview/negotiation phase for successful proposals. This negotiation phase may also involve further elaboration of project objectives, milestones and deliverables (for example, plans for the dissemination of results, or the likelihood of their uptake by industry or other users).

3.4.4 Project monitoring (scientific inquiry)

A major characteristic of the German approach to project management is the use of "Projektträger". These essentially are third party organisations, independent of both the sponsoring Ministry or research institute and the researchers, who undertake routine management of the project and ensure its smooth operation. This management function may include an element of monitoring. Regular feedback and interaction with the sponsors of the research are also prominent features of the system. It appears that Ministries and other agencies make use of a "pool" of Projektträger who have built up longstanding relationships with the Ministry officials. Many "Projektträger organisations are Government bodies in their own right, staffed by civil servants.

Significant emphasis is placed on the use of monitoring, which may be restricted to purely financial and administrative issues, or may extend to a more evaluative function and include the examination of industry interaction, dissemination activities and results uptake, for example. Monitoring may also be supplemented by periodic evaluations (see next section).

3.4.5 Evaluation of results, impacts and dissemination

Several methods are employed to determine achieved or achievable impacts (Kuhlmann, 2000). The most important are before/after comparisons, the control or comparison group approach, and qualitative analyses (which include plausibility checks and estimated judgements). These qualitative approaches may be operated alone, or combined with quantitative or semi-quantitative indicators, such as R&D expenditure, patents, economic, social and technical indicators, publications, citations, etc. Various data collection tools are also used, including existing statistics, questionnaires, interviews, case studies, panels, etc. Lastly, a range of data analysis tools may be employed, for example econometric models, cost/benefit analysis, other statistical methods, technometrics, bibliometrics, and peer reviews.

In general, most programmes undergo an evaluation, although there appear to be no systematic rules for this, even within ministries or ministerial departments. The use of a mid-term evaluation or review is fairly commonplace, as is a final evaluation. These often tend to be based on a mid-term or final report produced either by the project team or the project/programme manager. It is often the case that the Projekttäger, or other third-party body/committee undertakes the review. Evaluation reports prepared by third-party organisations are in turn reviewed by internal committees.

At first sight, it appears that there is a lack of emphasis on conducting evaluations for the purposes of ensuring that the appropriate researchers are conducting the research effectively. However, this objective is generally met through the periodic systemic evaluations which the German system appears to favour. Thus, rather than conducting numerous evaluations at the project or even programme level, researchers are evaluated at the institutional or supra-institutional level. Numerous examples of this are provided above.

3.4.6 Specific criteria

The range of specific criteria employed throughout the cycle of programme and project selection, assessment and management varies according to the nature of the research, its objectives, its participants, and intended outputs. Criteria are also tailored to suit the specific requirements of the evaluative task. Generally criteria are defined by the staff of the Ministry or Institution for which the research is being conducted, except in the case of external evaluations.

In general terms it is possible to identify a number of what may be termed "standard" criteria or indicators. These include: Track record of participants, Publications[31], Patents, inputs to regulations, policy inputs, technical reports, etc.

As an example of evaluation at the systemic level, the (scientific) criteria employed by the Wissenschaftsrat in its review of the Blue List institutes are:

- National and international integration of the institute in its main scientific field
- Coherence of the research planning and programme
- Qualified publications: e.g. number and quality of articles published in national and international refereed journals
- Total sum of external funding for research projects, especially peer reviewed funds
- Regular evaluation by a scientific advisory board
- Qualification and flexibility of the personnel
- Cooperation with universities and research institutes
- Joint appointments of leading academics with universities
- Participation of academics in university teaching and promotion of young academics
- Number of former academics of the institute who were appointed to professorships
- Number of academics who were invited to make presentations at important national and international conferences
- Number of academics who were invited for a research stay in other countries
- Number of external academics that were invited for a research stay in the institute.

As a footnote, however, it should be noted that the identification of possible lists of evaluative criteria that may be transferred to the Canadian system is a goal that must be viewed with caution. The specific national context within which these criteria are operated must be considered before any such attempts are made. Further, the scope of this study has precluded any form of benchmarking which could inform any such transfer. In terms of Government S&T, the risks are higher: as Kuhlmann (2000), citing Airaghi *et* al, 1999, argues, "the understandable desire for a standardised 'indicator tool-box' is not compatible with the tendency to pursue complex political goals".

3.4.7 Summary of conclusions

The German S&T system is characterised by a diversity of research actors. As a result, interactions between the various elements of the system are complex. Consequently, the provision of S&T research in support of policy goals is also complex, as are the mechanisms put in place to ensure the delivery of excellence. Nevertheless, it is possible to identify a number of general trends or features:

1. The programme definition process is very consensus-oriented. Extensive use is made of external scientific and technical advice, together with inputs from the eventual end-users of the research. The latter generally include industry, but the broader public may also play a role. Project definition is also highly influenced by external inputs.

[31] The Wissenschaftsrat does not accept that the use of the Science Citation Index is valid for all institutes of the Blue List.

2. The planning process, particularly at the project level, is strongly emphasised and involves a clear definition of milestones and objectives. The eventual use of results and plans for their dissemination may also be taken into account. This aspect is, in some ways, a prerequisite for the next feature.
3. Project management is frequently undertaken by third parties. These are often from specialist external management organisations or 'Projekttrager". Such managerial tasks range from relatively simple administrative and financial management to monitoring and evaluation.
4. Much emphasis is placed on project monitoring. This is may be conducted through a reporting system which typically will involve the preparation of mid-term and final project reviews.
5. The scientific and technical capability and capacity of research performers are assessed through two approaches: i) through open tendering and proposal processes which scrutinise the feasibility of intended projects and the ability of research teams to undertake them (possibly using peer review); and ii) through systemic evaluations of individual institutions or groups of institutions.
6. In some cases there is systematic evaluation of immediate research impacts. This is generally achieved by requesting the relevant information to be provided in the final project reports.
7. The emphasis on final reports (see previous point) tends to neglect any longer-term follow-up procedures aimed at identifying downstream impacts and effects.

3.5 Information sources

3.5.1 Persons contacted

Enzo Aufderheide	Wissenschaftsrat, Science Council
Prof. Dr Basler	BML, Federal Ministry of Food, Agriculture and Forestry
Rolf Busch	BML, Federal Ministry of Food, Agriculture and Forestry
Klaus Flath	BMWi Federal Ministry of Economics and Technology – Energy programmes
Hartmuth Grunau	BMBF, Federal Ministry for Education and Research – Production Technologies Programmes
Karl-Heinz Habig	BAM, Federal Institute for Materials Research and Testing
Peter Hart	BMU, Federal Ministry for the Environment
Bettina Hartwig	BML, Federal Ministry of Food, Agriculture and Forestry
Mr Herrman	BMU, Federal Ministry for the Environment
Dr Jaeger	PTB, Physikalisch-Technische Bundesanstalt
Stefan Kuhlmann	FhG – ISI, Fraunhofer Institute for Systems and Innovation Research
Horst Kunzmann	PTB, Physikalisch-Technische Bundesanstalt
Henk van Liempt	BMBF, Federal Ministry for Education and Research
Hans-Peter Lorenzen	BMWi, Federal Ministry of Economics and Technology
Dr Hubert Ottenwaelder	BMG, Federal Ministry of Health
Susanne Reichrath	Ministry of Education, Culture & Science, Saarland
Rainer Röhrig	BMBF, Federal Ministry for Education and Research – Laser Technologies Programmes
Dr Schött	BMBF, Federal Ministry for Education and Research – Key Technologies Programmes

Friedrich Tegelbekkers	Wissenschaftsrat, Science Council
Christian Uhlhorn	BMBF, Federal Ministry for Education and Research
Michael Welling	BML, Federal Ministry of Food, Agriculture and Forestry - Senat der Bundes-forschungsanstalten, Braunschweig

3.5.2 Bibliography

Airaghi, A., Busch, N.E., Georghiou, L., Kuhlmann, S., Ledoux, M.J., van Raan, A.F.J. and Viana Baptista, J. (1999) "Options and Limits for Assessing the Socio-Econoic Impact of European RTD Programmes". Report to the European Commission, DG XII, Evaluation Unit, Brussels/Luxembourg (Office for Official Publications of the European Communities). ISBN 92 828 3721 1.

Becher, G. and Kuhlmann, S. (eds.) "Evaluation of Technology Policy Programmes in Germany", Kluwer Academic Publishers, Boston/Dordrecht/London, 1995.

Belitz, H. and Lorenzen, H-P. "System Evaluation of the Promotion of Research Cooperation for Small and Medium-Sized Enterprises (SMEs) – the Concept of the Federal Ministry of Economics and Technology (BMWi)" US-European Workshop "Learning from Science and Technology Policy Evaluation", Bad Herrenalb,11-13 September 2000.

BMBF (Federal Ministry of Education and Research), Bundesbericht Forschung 2000, BMBF Publications, November 2000. Available at: http://www.bmbf.de/veroef01/ bufo2000/index.html#

BMWi (Federal Ministry of Economics and Technology), "The Courage to Innovate. Technology Policy – Paths to growth and Employment". Federal Ministry of Economics and Technology, December 1999.

Bührer, S. and Kuhlmann, S. (Eds.), "Evaluation of Science and Technology in the New Europe". Proceedings of an International Conference on 7 and 8 June 1999, Berlin. Bonn / Brussels: BMBF / European Commission, 1999, 209 S. (ISI-B-70-99).

Kuhlmann, S. "Evaluation of Research and Innovation Policies – An Overview from a German Perspective", draft paper for US-European Workshop "Learning from Science and Technology Policy Evaluation", Bad Herrenalb, 11-13 September 2000.

Kuhlmann, S. (2000) "Evaluation as a medium of science and technology policy: Recent developments in Germany and beyond" US-European Workshop "Learning from Science and Technology Policy Evaluation", Bad Herrenalb,11-13 September 2000.

Kuhlmann, S. "Moderation of Policy-making? Science and Technology Policy Evaluation beyond Impact Measurement: the Case of Germany", *Evaluation* Vol. 4 No. 2, 130-148, 1998b.

Kuhlmann, S. "German Government Department's Experience of RT&D Programme Evaluation and Methodology". *Scientometrics* 34 (1995), Nr.3, S.461-471 (ISI-P-102-95)

Kuhlmann, S. "Patterns of Science and Technology Policy Evaluation in Germany". *Research Evaluation* 5 (1995), Nr.1, S.23-33 (ISI-P-2-95).

Kuhlmann, S and Holland, D. 1995a "Evaluation von Technologiepolitik in Deutschland, Heidelberg, Physica-Verlag,1995.

Röbbecke, M. and Simon, D. "Evaluation of Leibniz Institutes. External Evaluation and Self-Evaluation" US-European Workshop "Learning from Science and Technology Policy Evaluation", Bad Herrenalb,11-13 September 2000.

Tegelbekkers, F. "Evaluation of the Blue List Institutes by the Science Council in Germany", Presentation at the OECD-Workshop on the evaluation of basic research, 21st April 1997, Paris.

Uhlhorn, C.D. "Evaluation of National Policies" – overview paper for US-European Workshop "Learning from Science and Technology Policy Evaluation", Bad Herrenalb, 11-13 September 2000.

3.5.2.1 Other documentation

BAM – Federal Institute for Materials Research and Testing, Annual Report, 1999.
"Projektorganisatiion Manual", BAM, BAM MP/G.2, 14 December 1999.
"Rahmenkonzept "Forschung für die Produktion von Morgen" (Framework plan "Research for the Production of Tomorrow"), Bundesministerium für Bildung und Forschung – BMBF, September 1999.
"Produktion2000+", RWTH, Aachen, 1999.
"Forschungsrahmenplan 1997–2000", (Research framework, 1997-2000) available from: http://www.bml.de.Forschungsrahmenplan/einf.htm
"Health Care in Germany", Federal Ministry of Health, available at: http://www.bmgesundheit.de/engl/health.htm
"Health Care Systems in Transition – Germany", European Observatory on Health Care Systems, available via: http://www.bmgesundheit.de/engl/
http://www.blk-bonn.de/papers/forschfoerd_summary.pdf
"The Wissenschaftsrat", available from http://www.wissenschaftsrat.de/wr_engl.htm
"Leitfaden für die Bewertung von Einrichtungen der Wissenschaftsgemeinschaft Gottfried Wilhelm Leibnitz (WGL) (Blaue Liste). Wisenschaftsrat, Geaschäftsstell March 1999.
"Stellungnahme zum Robert Koch-Institut", Wissenschaftsrat, December 1997
"Stellungnahme zum Paul-Ehrlich-Institut (PEI) – Bundesamt für Sera und Impfstoffe", Wissenschaftsrat, July 2000
"Stellungnahme zum Bundesinstitu zum Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin (BgVV)", Wisseschaftsrat, November 1999.

3.5.2.2 Internet sources:

BAM Profile – Federal Institute for Materials Research and Testing – http://www.bam.de
Bund-Länder Kommission - http://www.blk-bonn.de/
Federal Agricultural Research Institutes – http://www.dainet.de
Federal Ministry of Economics and Technology (BMWi) – http://www.bmwi.de
Federal Ministry for Education and Research (BMBF) – http://www.bmbf.de
Federal Ministry for the Environment (BMU) – http://www.bmu.de/english/fset1024.htm
Federal Ministry of Food, Agriculture and Forestry (BML) – http://www.bml.de/englisch/
Federal Ministry of Health (BMG) – http://www.bmgesundheit.de/engl/
Hermann von Helmholtz-Gemeinschaft – Association of National Research Centres – http://www.helmholtz.de
PTB – Physikalisch-Technische Bundesanstalt – Federal Institute for Physics and Technology – http://www.ptb.de

3.6 Annex BAM – Project assessment criteria (translation)

Effectiveness			
Position	Criterion	Criterion position	Points
1	Immediate us of results by public authority (federal offices, federal state authorities, judicial authorities, supranational authorities; e.g. procedures of legislation, licenses, chief expertises, or through legal defaults.)	N-1	3
2	High estimated economic benefits e.g. in key technologies, improved technologies, large broad-based effectiveness, public safety, protection of the environment, health care, etc.	N-2	3
3	High estimated benefits for self-management organisations of technology and economy (laboratories and control institutions, accreditation and certification institutions, professional associations, consumer protection associations, etc.)	N-3	2
4	High estimated benefits for one or more of the significant (or major) branches of the economy, and growing branches.	N-4	2
5	High estimated benefits for smaller branches of the economy or for single firms; estimated benefits for larger branches of the economy.	N-5	1
6	No concrete target group	N-6	0
Efficiency			
1	Estimated proportion of foreign funds x>30%	F-1	3
2	Estimated proportion of foreign funds 15% - 30%	F-2	2
3	Estimated proportion of foreign funds 0% - 15%	F-3	1
4	Estimated proportion of foreign funds 0%	F-4	0
5	Extensive and free of charge consulting service for authorities in written form (alternative to Pos.1-4)	F-5	2
Contribution to safety and reliability			
1	Direct relation to safety and environmental harmlessness of constructions, material and processes, including the development of control processes and the provision of reference materials and processes.	L-1	3
2	Direct relation to reliability of constructions, material and processes, including the development of control processes and the provision of reference materials and processes.	L-2	2
3	Indirect relation to safety, environmental harmlessness and reliability	L-3	1
4	No relation to safety, environmental harmlessness and reliability	L-4	0
5	Instructed variation	L-5	2
Impact on regulations and standards			
1	Essential cooperation during the establishment of legal regulations, legal mission, execution of legal regulations.	R-1	3
2	Pre-normative research; essential (leading) cooperation during the establishment of norms or other regulations.	R-2	2
3	Preparatory work and cooperation during the establishment of legal regulations, norms and other regulations.	R-3	1
4	No relation to the establishment of rules due to technical complexity or due to peculiarity of the question.	R-4	1
5	No relation to legal regulations, norms and other regulations.	R-5	0
Subsidiarity			
1	Directive by the federal offices, BAM as solely authorised instance mandated by legislator.	S-1	3
2	BAM beneath other mandated instance; BAM has specifically high competence as unique selling point or has to build high competency due to strategic reasons; other do not notice the task.	S-2	2
3	BAM has specifically high competence.	S-3	1
4	Several instances with equal competence	S-4	0

BAM Project evaluation schema

Project Number: Date:

Criteria	Criterion Position	Points (0-3)
Criteria for effectiveness and efficiency		
Effectiveness		
Efficiency	F.	
3-criteria filter		
Contribution to safety and reliability		
Impact on standards and regulations	R.	
Subsidiarity	S.	
	Summarised points:	

Remarks:

Evaluation:

	Highest priority
	High priority
	Medium priority
	Less priority

4 European Study Report: Sweden

Prepared by Steven Glynn

4.1 Introduction

Following the remit of the terms of reference for the study, this element of the review focused on relevant activities in Sweden. This report first presents a very brief overview of the Swedish research system, providing context to the discussion of evaluation and assessment processes that follows. Section 4.2 discusses the difficulties of presenting a general overview of evaluation in Sweden. Leading on from this, section 4.3 is concerned with specific examples, highlighted assessment and evaluation procedures for a range of organisations. The report concludes with a discussion of whether any general trends can be identified.

4.2 Main Findings

4.2.1 Overview of system

The Swedish system of Government differs from that found in many other countries. In Sweden the ten Ministries (plus the Prime Minister's Office) that make up the Government office are very small. In place of larger ministries, there are a number of National Boards and Agencies. These Boards and Agencies are independent of the Ministries, although they may report back to them, and a number of them can decide on regulation in their specific areas. One example of this is the Environment Protection Agency. The EPA is much larger than the Environment Ministry – with around 500 employees compared to around 100 in the Ministry – and the main core of expertise tends to lie within the Agency. While the EPA is responsible to the Ministry, and receives (very) broad instructions as to how it should allocate its funds, it is separate from the Ministry.

Currently there are several research councils under the Ministry of Education and Science which are responsible for funding basic research. There are also a number of more sector specific research councils that fund activities of a more applied nature. However, perhaps the most important feature to understand about the research council system is that it is changing. As of January 1st 2001 all the separate research councils will be integrated into one large Science Council. One major rationale for this change was a perception that the operation of the old research councils had become too routine and that the distribution of funds had not adapted sufficiently to reflect changing conditions and requirements. However, at the time of writing, there seemed to be much uncertainty as to how the new system will operate.

The Swedish system is further complicated by the existence of private research foundations. These foundations were established in 1994 by the then conservative government with money derived from the former Employees Investment Funds. In many ways these private foundations can be seen as mirroring the research council system but in the private sphere. In fact, the foundations actually disburse more money than the research councils do. However, although they are private, the government nevertheless appoints the boards of these foundations.

Given this rather diffuse nature of research in Sweden it is difficult to assess where the 'government' is involved in evaluating research and ensuring excellence. Thus, rather than a

centralised system, responsibility for these activities appears to be largely devolved to the individual organisations concerned with performing research.

4.2.2 Overview of evaluation

There are no general overarching guidelines and/or routines for achieving and monitoring excellence in the Swedish system. Quality assessment is undertaken within the different research councils, agencies and foundations making it difficult to make any general statements about a 'Swedish' model for ensuring research quality. It is therefore perhaps most beneficial to discuss particular examples, which will be done in the following section, before drawing some general conclusions.

4.3 Specific cases

This section concentrates on the processes for assessment and evaluation in a range of organisations within the Swedish research system. Given the limitations of time and resources it was not possible to comprehensively cover the entire research system.

4.3.1 The Swedish Environmental Protection Agency - EPA

The Environmental Protection Agency's most important tasks are:

- The promotion of ecologically sustainable development, contributing to achieving the objectives by taking on the role as co-ordinator and driving force in environmental work both nationally and internationally;
- To compile and disseminate knowledge about the environment;
- To promote sustainability in trade and industry, products, waste and infrastructure in co-operation with the sector authorities as well as local and regional authorities;
- To contribute to the realisation of environmental policy and the achievement of established goals; and
- To follow up and evaluate the condition of the environment and the work done to form the basis for further development of environmental policy.

The Agency has received a Government research grant since the 1960s and has been the major funder and financial partner to universities in the environmental area. As mentioned previously, in 1994 private research foundations were established. Amongst these was MISTRA, the Swedish Foundation for Strategic Environment Research. At first. the EPA undertook collaborative work with MISTRA, although MISTRA tended to perform research regarding abatement strategies while the agency tended to focus on the origins of environmental problems. Later, in the 1990s, there were severe cutbacks (generally in funding levels) and the Agency's research grant was unpaid for two years. At this time MISTRA took over responsibility for environment research. In 1999 the research grant was given back to the Agency but at a lower level than previously.

The Agency currently spends around £6 million on research and this is focused on more short-term issues. Under the reorganised research system (as discussed earlier), MISTRA will continue to exist but there will be a new research foundation that will concentrate on agriculture, forest, environment and planning issues, and part of the Agency's duties will be in this organisation. This new organisation will be the biggest in Sweden in relation to environment research.

4.3.1.1 Broad evaluations of the area

At the very broad level, the EPA has initiated an evaluation of environmental research in Sweden - *R&D for a better environment*. This takes place every 4 years and examines the current direction of research, priorities for funding in the future, and identifies where the responsibilities for research should lie. Although this report is, due to its broad focus, more of a discussion paper, the Government tends to refer to the document when seeking advice on funding allocations. It is thought that the new foundation will take on the responsibility for this document. The initiative has been led by the EPA, which sets up a working group after inviting contributions from other groups and organisations. However, the process is rather informal.

The EPA is also involved in relevant general evaluations that are performed on specific areas of environmental research (see below). These involve various scientific and user bodies and an international panel. These general evaluations tend to largely focus on the scientific issues and research quality although the Agency does try to introduce questions of relevance. The Agency also undertakes independent reviews, for example of wildlife protection funding, and tends to have a close relationship with the users of environmental research funding.

4.3.1.2 EPA funded research

Regarding money spent in the agency, there are two types of financing. The first covers programme work. The decision to focus on particular areas (i.e. programme definition) results from a consultative process wherein relevant groups, including NGOs, individual researchers, etc., are invited to contribute. A programme group (of 5-7 people) is then established. Over the course of a year this group outlines a programme. The board then decides whether to finance it. Around 30 per cent of the programme money is reserved for the project group, the rest is open to annual research applications. Within the agency the project group judges these applications in terms of their scientific quality. Programmes generally have a 4-5 year life span. When programmes come to an end, a report is produced which reviews the whole programme, looking at both the relevance and quality of the work done. Other aspects of the work will also be reviewed, such as its contribution to education, whether researchers have served on expert groups, etc.

The EPA also has five scientific committees (covering land/air, aquatic, environmental toxicology and medicine, socio-economic aspect, and wildlife fund). Each of these committees has a budget of around £1 million which is used to fund projects (of between one to three years' duration. Part of the contract for these research projects stipulates that the supported researchers must report back, giving an overall review of the science conducted together with a more 'popular' summary concerning the potential and actual role of the scientific information produced.

The research contracts also allow the Agency to have access to research findings during the course of the projects. If any (environmental) issue arises, the Agency will contact researchers and request from them an overview of the area and the present state of knowledge. Thus, although the government tends to adhere to its strategic plans, when issues do arise it may request special enquiries related to specific questions. This relates closely to the Agency's role in assessing the impacts of research.

On an ongoing basis, the EPA conducts interviews with local/regional authorities, industry, other Ministries etc., to check that the research it is conducting is satisfactory to their needs, and whether they have made use of it. Related to this is the Agency's role in the

dissemination of research - ensuring that the knowledge produced by researchers is used wherever possible. While the dissemination of scientific knowledge is largely taken care of by publications in scientific journals, attention is also given to wider aspects of dissemination.

There are two aspects to this latter issue. The first is 'sender led', where individual researchers or research groups send out information. The other is receiver initiated. An example of this latter case is illustrated by a recent situation: Sweden has invested a lot of money to help ensure that regional authorities' activities are environmentally sound. To get this funding, the regional authorities prepare applications, but the EPA assists in the process by identifying appropriate work they could undertake. In one particular case, the EPA focused on a type of glue used in construction. The glue contained high levels of PCBs, and the Agency was already sponsoring a project looking at the leakage rates of these chemicals. While the research had not been finalised the Agency asked the research group for a knowledge synthesis of the results so far and the regional authorities then used this information. This provides an example of research that was not finalised but which was nonetheless useful and prescient.

4.3.2 The Natural Science Research Council (NFR)[32]

There are a number of levels at which evaluations are performed in the NFR, from individual projects up to whole areas of research.

4.3.2.1 Individual projects

All project proposals are peer reviewed and then passed to the council for approval. The council works with programme committees and there is a standard form for expressing judgement on each application. Researchers usually receive grants for three years. At the end of the grant period, researchers are required to report back to the council so that their work can be evaluated. This evaluation focuses on the 'quality of the science' that has been done. Evaluations consider various criteria including publications and researchers' standing in the international arena, numbers of students supervised, numbers of PhDs produced, etc. However, it is the standing of the science which forms the most important criterion and if a researcher has published extensively in good international journals, then the other categories (such as student numbers) will be viewed with less importance.

4.3.2.2 Research Programmes

Over the last six years the council has had a number of research programmes in operation. The main programmes concern research positions/appointments but these are not the concern this study. The programmes are proposed by the subject committees in the council, highlighting areas for which they feel there is a need for support. These have not been particularly popular with the research community, which views them as a disturbance in the system. In principle the council should evaluate these programmes but this has not been done.

4.3.2.3 Research Areas

Every five to six years much larger evaluations are undertaken into particular scientific areas (for example, atomic physics). The relevant committees of the council identify the specific areas which are to form the subjects for evaluation. There is not normally any opposition to the way the areas are divided. These area evaluations are conducted by a team of five or six

[32] The NFR ceased to exist as of 1 January 2001, as a result of the reorganisation of the Swedish research system.

international experts, who are provided with specific guidelines. The purpose of these evaluations is to inform the Research Council about the scientific quality of research projects as seen in the international context. While the projects form the main unit of analysis and basis for the evaluation, the expert group is invited to comment on any relevant structural issues. In addition to evaluating the work that has been done, the reports produced by the expert groups are expected to be forward looking and strategic.

The evaluations are based on information submitted by the principal project grant holders. Generally, this information is expected to cover[33]:

- A short description of the objectives of the project;
- A summary of the results of the project during the last five years regarding scientific results, new methods, new development of equipment, applications etc.;
- A summary of the resources (from both NFR and other sources) available to the project during the last two fiscal years;
- A prognosis on the future development of the project in terms of objectives, resources, time schedules etc.;
- A complete list of publications for the last five years relating to the project with the most important publications attached;
- A summary of examinations, dissertations for the degree of PhD etc. produced within the project during the last five years;
- A short CV for the principal grant holder; and
- A summary of those current activities of the principal grant holder that fall outside the scope of the project of the present review.

The extra work this entails means that it is not popular with the research community but there is generally understanding of what is being done and many researchers see it as an opportunity to possibly extract more money for their particular area.

In addition to reviewing the written material, the expert evaluation group is invited to carry out site visits to interview the research groups where the project are being undertaken.

In reviewing each project the evaluation group is expected to comment on

- The scientific quality of the results obtained;
- The scientific value of proposed projects (including the question of possible improvements by changing the aim and direction of the project);
- The merits of the methods used and proposed;
- The capability of the project leader and the staff (including issues such as size and composition of the group);
- The adequacy of existing and proposed research positions, facilities and equipment;
- Other considerations or viewpoints which may be of importance for the projects; and
- The question of increased, unchanged, or decreased support. Or the termination of projects (with constructive alternatives where possible).

At the end of the evaluation the expert group is expected to produce a report that covers both the general state of the research area in Sweden and the individual projects themselves. With

[33] From NFR Minutes, No.116, *Evaluation of Research Projects*, February 3 1993

regards to the status of the research area in Sweden, the report should include discussion of the general scientific level of the research performed, encompassing any structural or organisational problems. The group is also asked to comment on the need for any expensive equipment and any other points of general relevance. Regarding the individual projects, the group is expected to produce a page on each project with specific comments and recommendations (recommends, strongly recommends, most strongly recommends), along with a ranking for the project. The ranking system used is:

Excellent	- Research at a very high international level; of great international interest with a broad impact and with publications in internationally leading journals; the researchers are among the leaders in the field
Very good	- Research at a high international level; of international interest with impact within its sub-field and with publications in internationally leading journals; the researchers are among the leaders in the sub-field
Good	- Research at a good international level with publications in internationally well-known journals; the researchers have a good international reputation within the sub-field
Fair	- Research that only partly is of good international standard and only partially published in well-known international journals
Poor	- Research of insufficient quality[34]

Research groups that are not highly ranked have been known to complain, arguing, for example, that the experts were not sufficiently expert in their area. Any such complaints are noted (if they are in written form than they are distributed to the members of the Research Council and to the Programme Committee) but no further action is taken and the council does not engage in any further dialogue with the experts once they have delivered their report. It is assumed that procedural issues highlighted by valid complaints may be modified in the conduct of future evaluations.

Once delivered, the report is distributed to the members of the relevant Programme Committee, to all grant holders concerned and, on request, to any other agencies e.g. universities, ministries etc., or persons who express an interest. (All reports of this kind become, under Swedish law, public documents). The reports of the international experts are used extensively by the council as part of the assessment process of new applications. Thus groups that have been judged "excellent" may receive more money while those less highly ranked may find the amount of money they receive cut – although this is not necessarily always the case. In judging applications previous grant reports are also used.

International evaluations have been used in Sweden since 1977 and the format has remained relatively unchanged. However there are a number of problems with this system. One particular drawback is that the experts are peers from within the same field and nearly always tend to defend the field. It is unlikely that they would argue that the field was not relevant – it is always a question of the need for more support. As a result, although the reports can be used to redistribute money within areas, they tend to be relatively ineffective for the redistribution of funds between areas. Whilst this is a widely recognised problem, there does

[34] The categories used in this system bear some resemblance to that used in the UK by the Higher Education Funding Council for England in its four to five yearly Research Assessment Exercise (RAE) of university departments. The RAE, however, uses a combination of historical data (publications, etc.), current data (student numbers) and future intentions (departmental strategy reports) for the allocation of funding. Also, its is not concerned with Government funded research as defined for the purposes of this study.

not seem to be any good mechanism for circumventing it. (Moreover, it is a problem which is not restricted to Sweden alone). The council has attempted to carry out an evaluation of an entire subject area (physics), again using international experts. However, the sheer size of the field led to difficulties with the evaluation and it was unable to produce detailed results.

Money can be redistributed using the special programmes discussed earlier. To support these special programmes, funds are drawn from various areas – however, as was mentioned, this has not been a popular practice. Changes in funding allocations within the system also stem from specific Government policy actions. For example, a recent Government Bill allocated more money for the system as a whole but the additional funding has been targeted towards specific areas such as IT and biotechnology/genomics. It is only through this general allocation of funding that the Government exerts any direction over the council, apart from general statements about requirements to support research excellence.

It is also worth noting that an international team of experts has been used to undertake an evaluation of the working methods of the council. This resulted in a few minor operational changes.

4.3.3 The research foundations

The following sections discuss assessment and evaluation of research in two of the private research foundations that form an important part of the Swedish research system.

4.3.3.1 The Swedish Foundation for Strategic Research (SSF)

Under the terms of its statutes, the purpose of the Foundation, established in 1994, is to promote the development of research environments that are of the highest international class and which are seen as of importance for the development of Sweden's future competitiveness. Strategic research is seen as research that is judged to be of long-term benefit to Sweden. The benefits are seen to involve one or more of the following aspects:

- The power of internationally first-class research environments to attract unique competence and international investment to Sweden;
- People with graduate degrees who through broadened and improved research training are attractive for appointments primarily within industry and the administration, but also at universities;
- Research results that may form the basis for the development of existing or new enterprises;
- Increased quality of life through more employment opportunities, improved working conditions and improved health; and
- Research that acts as a focus for international co-operation through which knowledge of relevance to Swedish industry may be brought into the country.

The Foundation's scientific policy is outlined by the executive director and the four strategic advisory committees – the annual strategy document that they produce forms the basis for the Foundation's call for proposals. Programme proposals are assessed for scientific quality – by at least three internationally prominent experts, and for relevance – by organisations such as the Federation of Swedish Industries or the Swedish Association of Local Authorities. Grantees are required to submit an annual report to the foundation.

Once a year a person from the Foundation's secretariat participates at a programme board meeting and, in addition, the programmes are evaluated half-way through. As with the

applications procedure, international experts are used and their report is used to determine the programme's direction for its remaining duration. After four years the programme board is required to deliver a report which forms the basis of its final report. It is an expectation that research results will be published in recognised journals.

4.3.3.2 The Swedish Foundation for Strategic Environmental Research (MISTRA)

As with SSF the purpose of MISTRA is to support the development of international class research environments that are important for Sweden's long-term competitiveness. MISTRA supports strategic environmental research - that is, research with a long-term perspective directed towards solving major environmental problems – which:

- meets the highest scientific standards;
- is guided by a vision of a more environmentally sound society;
- has as its aim system changes promoting sustainable use of resources, radically reduced environmental impacts, or essential new knowledge about environmental problems and their relative significance;
- contributes to Sweden's competitiveness; and
- has clear and measurable objectives, enabling it to be effectively evaluated.

Since it was set up, in 1994, some 200 outline ideas have been submitted to MISTRA. From these about 40 programme proposals were developed leading to the 22 programmes that the foundation currently runs. As with the SSF, programme proposals are assessed for both scientific quality and relevance. Again, international experts are used to assess the scientific quality. MISTRA has had no problems with these assessments – there has been no difficulty in engaging highly qualified experts and their suggestions have been very useful, giving the board a good basis on which to form decisions. In its examination of relevance in the Swedish context, MISTRA invites comments from industrial sectors and government agencies that are seen as having an interest in the research. However, this process has been more problematic. Most responses have been favourable but fail to offer any analysis of what the critical knowledge gaps are from the users' point of view.

MISTRA also evaluates the utility of the research conducted on completion of the first phase of a programme. This is seen as important given that programmes are geared towards finding solutions. However, MISTRA officials could not identify any good models for evaluating the types of programmes they support. The problem is that MISTRA provides long-term funding and aims for radical environmental improvements, making it difficult to assess utility after only a few years. Therefore the Foundation concentrates on potential utility – 'will the results of a successful programme help to solve the problems identified?' Currently these assessments are undertaken by the executive committees of the programmes but ultimately the Foundation would like to have this performed by more independent experts.

4.4 Overall conclusions

As mentioned previously there are no general overarching guidelines and/or routines for achieving and monitoring excellence in the Swedish system. Quality assessment is done within the different research councils, agencies and foundations thus making it difficult to make any general statements about a 'Swedish model' for ensuring research quality. In addition, the 'diffuse' nature of the Swedish research system makes it difficult to assess the role of the Government (as a centralised entity). As noted, the agencies (executive bodies) are

larger than ministries and operate independently, although they do report to their relevant ministry. The only guidance that the agencies receive from the ministries tends to be very broad (for example, that the Agency should support work relevant to its remit). The Research Councils receive funding from the Government but, while some may be earmarked for specific areas, there is little dialogue between the Councils and the Ministry of Education and Sciences which distributes the money. Again, Ministerial guidelines are very broad – i.e. to support research excellence. Meanwhile, the Foundations, while having boards appointed by the Government, are private concerns.

In terms of evaluation, three points seem to stand out:

- Firstly, there seems to be a strong reliance on the use of international experts to assess the quality of science, irrespective of the organisation. This is perhaps understandable in a country like Sweden where the research community is relatively small.
- Secondly, peer review is widely used to assess project proposals. In all cases that were examined, some form of peer review was employed in the *ex ante* assessment of project ideas.
- Thirdly, while the research councils' priority is scientific quality, the agencies and foundations will also evaluate programmes on the basis of their relevance to potential users of the research in Sweden.

4.5 Information sources

4.5.1 Interviews conducted:

Erik Fellnius, Research secretariat, Swedish Environmental Protection Agency

Lars Gidefeldt, Deputy Secretary General, Management Secretariat for Co-ordination and Planning Natural Science Research Council (NFR)

Mats Johnsson, Ministry of Education

Others contacted:

Gunnar Leman, Deputy Director, Management Secretariat for Co-ordination and Planning Natural Science Research Council (NFR)

Mats Benner, Department of Sociology, Lund University

Ulf Sandstrom, SISTER, Stockholm

4.5.2 Other sources of information:

MISTRA, 1999, *Annual Report*
NFR, 1993, *Evaluation of research projects*, NFR Minutes no.116
Swedish Foundation for Strategic Research (SSF): http://www.stratresearch.se/eindex.htm
Swedish Foundation for Strategic Environmental Research (MISTRA): http://www.mistra-research.se/
Swedish Environmental Protection Agency (EPA): http://www.environ.se

Swedish Natural Science Research Council (NFR): http://www.nfr.se

5 European Study Report: United Kingdom

prepared by Philip Hills

5.1 Introduction

The science and technology we are asked to address is that undertaken, commissioned or otherwise acquired by government departments to inform and support policy decisions and activities. It includes research and scientific advice as well as the collation, storage and dissemination of information. It covers the full range from basic research and the natural sciences to applied research and the social sciences. For the purposes of this paper we refer to this complex of work as 'government science and technology'.

We examine UK government science and technology in terms of three broad sets of institutions. First, the higher education system can be seen, with minor qualifications, as separate from the other relevant activities. Secondly, three departments, the Treasury, the Cabinet Office and the Office of Science and Technology, impart an element of direction and coordination. Thirdly, the remaining departments involved operate largely independently within a rather loose framework of principles. This paper treats these institutional sets in turn and then draws together the resultant findings.

5.2 The Higher Education System

The higher education system is related only rather indirectly to the theme of this paper. Much of the scientific expertise on which departments draw comes from academia. Much of the research that government departments commission contributes both to policy formulation and to the science base generally. Universities, like government, are concerned to ensure excellence. They do so primarily through the peer review process as applied through reputable academic journals. Government, in its turn, informs policy towards universities – specifically in the context of allocating research funds – through a quinquennial Research Assessment Exercise (RAE) that is itself heavily based on peer review and the assessment of publications. Indeed, a problem that some foresee is that the volume of peer review activity arising both within the higher education system and, increasingly, from government may become self-defeating. A few of those to whom we have spoken, with experience of both using and acting as peers, cited cases where those asked to act in this capacity refused on grounds of overload and claimed to have done so themselves. One also attributed a perceived diminution in peer review quality to the same cause.

5.3 The Central Departments

The 'central' departments of UK government in this context are the Treasury, the Cabinet Office and the Office of Science and Technology (OST). The last named is formally a part of the Department of Trade and Industry (DTI). These departments set a general framework for policy through various mechanisms that affect government science and technology, but which do not prescribe in detail how it is to be managed. Policy at the more detailed level is almost entirely devolved to the individual government departments directly concerned.

The Treasury exercises influence mainly through the three yearly Comprehensive Spending Reviews and their annual up-dates. These determine the amount of money available to departments in total and by broad categories. The allocation each department secures for science and technology is clearly likely to reflect wider national priorities. The allocations are supplemented by Public Spending Agreements (PSAs) between the Treasury and each department which consist of a negotiated list of objectives that the department pursues in return for the money which it receives. Each PSA is underpinned by a Service Delivery Agreement which fills in some (but not much) detail about the way in which the department seeks to achieve the objectives. In a broad sense the process sets a framework for S&T activity, but rarely in any specific way.

The Treasury also issues what is known as '*The Green Book*'[35]. This contains very detailed and often technical guidance on *ex ante* appraisal and *ex post* evaluation[36]. It covers such techniques as cost benefit analysis, including handling risk and uncertainty, the cost of capital, discounting, rates of return and so on. It also discusses managing evaluation and provides a checklist of necessary actions. In practice, however, much of the detailed advice is only applicable to very large undertakings such as building a hospital or a major road. It would be too cumbersome for any but an exceptionally large research project. The advice on managing an evaluation is in fairly general terms and is very much in line with practice in, for example, the DTI (see below). The Treasury, in consultation with OST and the Cabinet Office, are currently in process of revising *The Green Book*.

The Office of Science and Technology does not intervene extensively in the detail of departments' S&T activity, but confines itself to influencing strategy and the 'key principles applying to the development and presentation of scientific advice for policy making'. Guidance on these key principles was first issued in 1997.[37] Two reports have been published on the implementation of this document, which was seen as not quite clearly enough focused[38]. This led to revised guidance published in July 2000[39] which stresses the importance of identifying issues on which scientific advice or research may be needed as early as possible, acquiring the best expertise, asking the right questions and ensuring transparency. This reflects the emphasis on what is, essentially, peer review although it is mediated in a variety of forms.

There is also a Ministerial Science Group (MSG) under the chairmanship of the Minister for Science[40], which requires all relevant departments to draw up S&T strategies in consultation with OST. This group consists of middle ranking or junior ministers (including those from the devolved administrations in Scotland, Wales and Northern Ireland) which, although it may have less authority than a full Cabinet Committee concentrates more attention on the

---

[35] HM Treasury (1997) *'The Green Book': Appraisal and Evaluation in Central Government.* Second Edition

[36] In UK government the usual practice is to refer to *ex ante* evaluation as 'appraisal' and to reserve the term 'evaluation' for the *ex post* activity. The term assessment is often used to cover the whole spectrum from appraisal, through monitoring (during programme or project life) to evaluation.

[37] Office of Science and Technology (1997), *The Use of Scientific Advice in Policy Making: A Note by the Chief Scientific Adviser.*

[38] The second of these reports – Office of Science and Technology (1999), *The Use of Scientific Advice in Policy Making: Implementation of the Guidelines* provides some information, now slightly dated, on the activities of individual government departments additional to that in the present paper.

[39] Office of Science and Technology (2000), *Guidelines 2000: Scientific Advice and Policy Making.*

[40] The Minister for Science is a middle ranking minister reporting to the Secretary of State for Trade and Industry

issues. In general, the approach adopted consists of involving as much scientific advice as possible and integrating it with policy more closely than before. This is done through, for example, the Council for Scientific and Technology (CST), which is chaired by the Secretary of State for Trade and Industry, the cabinet minister responsible for science and technology. In the course of the relevant consultations the MSG and OST will take into account overall priorities as indicated, for example, by the on going Foresight activities.

The Cabinet Office have recently taken significant steps to improve policy making across government. These started with the White Paper *Modernising Government*[41] published in March 1999 and were followed by more detailed documents, the most significant of which, in the present context are *Professional Policy Making for the Twenty First Century*[42] and, to some extent, *Adding It Up*[43]. These emphasise the importance of 'evidence based policy making' and of evaluation. They do not go into any detail on methodology. The Cabinet Office, however, intends to set up a 'centre of excellence' on evaluation within its Performance and Innovation Unit. It does not envisage this centre undertaking specific evaluations, but sees it as encouraging best practice and advising departments when appropriate. While it is not planned that it should have prescriptive powers, it is likely to be influential. Departments that reject its advice will need to have good arguments for having done so when they seek funds from the Treasury to continue the policies concerned.

Central departments do not audit or oversee other departments' arrangements, though the National Audit Office[44] (NAO) can always exercise the option to do so. In practice, however, NAO investigations are most likely to pick up on details of scientific assessment only as a by-product of a more general investigation. Nor, incidentally, do any of the departments to whom we have spoken carry out regular formal reviews of the assessment procedures. More informally, arrangements are reconsidered and, from time to time adjusted, as experience dictates.

In summary, the central departments wish to ensure that, wherever it is appropriate, policy is underpinned by sound science. They are not much concerned in this context whether that underpinning derives from the existing knowledge and expertise of advisers to whom they have access or from specially commissioned research. While these departments influence and offer guidance on broad priorities, strategy and general principles, they do not normally involve themselves, on a day-to-day basis, in the detail of the way 'executive' departments handle their scientific arrangements[45].

For completeness it should be added that, while the European Commission has its own evaluation arrangements, these do not impact directly on the ways in which individual governments seek to ensure excellence.

---

[41] Cabinet Office (1999) *Modernising Government,* Cmnd. 4310 Stationery Office
[42] Cabinet Office (1999) *Professional Policy Making for the Twenty First Century*
[43] Cabinet Office (2000) *Adding it up – Improving Analysis and Modelling in Central Government*
[44] The National Audit Office are, however, responsible to Parliament, not to the government
[45] These generally decentralising tendencies are complemented by measures to improve interdepartmental cooperation on cross-cutting policies. They do not, however, involve central department intervention except as a last resort. For more detail see: Cabinet Office (2000), *Wiring it up: Whitehall's Management of Cross-cutting Policies and Services.*

5.4 Assessment Systems

In 1989 the Cabinet Office (then incorporating part of what is now the OST) issued quite detailed guidelines on R&D Assessment (including everything from pre-project/programme appraisal, through managing and monitoring to *ex post* evaluation)[46]. This guidance has not been formally withdrawn, but it is recognised (although this has not been set out explicitly) that the circumstances and purposes of different departments vary to an extent that makes detailed uniform guidance impracticable in application. It is, therefore, unlikely to gain acceptance and difficult to enforce.

The current view is that a more flexible system is more likely to ensure that no important issues get overlooked; that departments focus the issues well; and that they tap the best and most appropriate available scientific resources. Such an approach is based on the view that the best guarantee of quality is to enlist as wide a range of scientific opinion as possible through open consultative processes, which encourage all relevant interests to come forward and all sources of expertise to identify themselves. The advent of the internet has greatly facilitated this process. At later stages peer review is incorporated, wherever appropriate. This, more than formerly, open approach seems, to some extent, to be a reaction to the perception that where problems have arisen (as, for example, in the BSE crisis) they have been partly due to an excessive tendency to secrecy.

Nevertheless, while the detail of the 1989 Cabinet Office guidance has fallen into disuse, many of its principles have become fairly widely accepted. It promoted an integrated approach to managing programmes on a basis pioneered and first applied in the Department of Trade and Industry (DTI). Under this arrangement every programme is approved, targeted, monitored and finally evaluated on the basis of a ROAME (Rationale, Objectives, Appraisal, Monitoring and Evaluation) statement.

The rationale section of this statement sets out the overall justification for the programme and its basic aims. From these grow a hierarchy of objectives for the programme as a whole and thence leading to the objectives, targets and milestones of individual projects. As the programme progresses, these become the basis for monitoring (and adjustment as necessary), appraising new constituent projects and, finally, evaluation. The sections of the statement relating to these activities also specify timing and procedures.

The main impacts of the ROAME approach in its various manifestations have generally been positive in that it has, first, focused activities more closely on clearly defined objectives. Secondly, it has given more impetus to an evaluation culture – especially *ex post* evaluation with resultant feedback into new policies and programmes: that is systematic learning from experience.

Some aspects, however, have been found to be negative in that it easily lends itself to rigidity and undue bureaucracy. As these problems have become manifest the system has evolved and continues to do so. It is more explicitly recognised that the same arrangements cannot be applicable to all cases.

---

[46] Cabinet Office (1989), *R & D Assessment: A Guide for Customers and Managers of Research and Development.*

Objective setting has sometimes been seen to be distortionary. For example, insistence on verifiability can trivialise and result in over concentration on things that could be easily quantified, while other - sometimes more important - aspects got ignored. There has sometimes been too much concentration on short, rather than longer term, objectives because they were easier to think about. And it is sometimes forgotten that any set of objectives is, at best, only a partial description of what the activity is about. If the operators of policies and programmes concentrate solely on the formal achievement of objectives they are liable to distort the overall balance of the programme - especially if the objectives were less than perfect in the first place. Some would argue that this has happened in education and to some extent in the health area. Finally, there are complaints – probably sometimes, but not always, justified - that the whole exercise can become excessively time consuming.

5.4.1 The Department of Trade and Industry (excluding the OST)

The DTI, excluding the OST, is the department responsible for promoting competitiveness, enterprise and innovation in industry thus contributing to economic development and the quality of life. It funds around £300m of research annually. Much of this is applied research in support of industry. Some is to maintain standards (e.g. of length and mass) and some is in pursuit of the national space programme, largely through the European Space Agency. Almost all of it is contracted out.

In the DTI the ROAME system continues to operate, fundamentally in the form described in the previous section. Problems have been identified, however, and these have led to changes of detail and to a current review in relation to *ex post* evaluation. Some of these problems relate to the criteria for initiating new programmes which have not always been as useful, in practice, as they might have been because they have concentrated too heavily on theoretical conceptions of market failure. The emphasis on the programme as the unit of assessment has also been found sometimes to result in too narrow a focus to optimise policy guidance. It may be that in time this will result in ROAME statements becoming more like business plans, and evaluation covering whole policy areas rather than individual programmes.

In the DTI officials or specifically contracted consultants monitor programmes and projects internally.

Ex post evaluation is the responsibility of a committee of officials chaired by a senior finance officer and including policy makers and evaluators. This committee determines a programme of evaluation work and ensures that all major activities are regularly covered. A sub-group provides the forum in which policy makers, evaluators and the evaluated discuss the completed reports. It acts both as a quality control on evaluation work and to ensure that recommendations which emerge are taken into account by those concerned.

Most of the actual evaluation work is undertaken internally. In the case of science and technology there is a separate unit, detached from those who run programmes or projects, with a culture and recognised tradition of independence. It undertakes most DTI evaluations itself, though in a minority of cases the unit employs external evaluators. Because it is divorced from the activities immediately under scrutiny, neither the unit, nor outside contractors are under any obligation to those they are evaluating. The resulting work, therefore, has a degree of, but arguably not complete, independence. The unit has the advantage of unfettered access and a relative insider's understanding of the context.

Evaluation is primarily focused at the programme level. There is a view, however, that separating programmes from associated policies diminishes the input to decision making. In addition it has proved difficult to compare programme evaluations so as to inform future choices. These considerations prompt consideration of broadening the units of assessment and, possibly, standardising evaluation criteria to facilitate comparison. These two possibilities conflict, to some extent however, in that the first would involve wider and more varied research, while the second would tend to reduce variation. The way in which the system may evolve is, therefore, still uncertain.

Evaluators typically start from a consideration of the original plan for the programme, normally drawn from its ROAME Statement. They assess the extent to which the justification set out in the rationale has been vindicated by experience and whether the projects that have been supported appear to be consistent with the stated objectives. They draw on monitoring data acquired during the programme, making allowance for the focus of the programme having evolved over time in accordance with the logic of its own progress, provided such evolution is conscious and not simply drift. From the files, they form a view of the effectiveness of programme management. They discuss what they find and the related issues with the officials who managed the programme.

The next phase of the evaluation usually includes interviewing representatives of participating organisations face to face and/or by telephone. Sometimes these are supplemented with written questionnaires. Among such interviewees are often included a sample of those whose applications to participate were rejected and sometimes evaluators also contact firms or universities who might have been expected to participate but did not. Often, evaluators discuss the programme with outside experts (peers) who are knowledgeable about it or its subject matter. They may carry out, or commission, a bibliometric study of the resulting publications, especially where more fundamental research is involved and, in a few cases, other outputs such as patent applications or registrations have been quantified. This last, however, has not played more than a very subsidiary role in evaluation. DTI find that this combination of peer and user opinion with objective indicators and statistical data enables evaluators to form an overall view of quality and the extent to which original objectives, were appropriate, and have been achieved.

Most evaluation is aimed at assessing immediate results and impacts because this is usually necessary to inform decisions about future activities. It is recognised that long term impacts are too delayed to inform relevant policy decisions. Moreover, evaluation long after completion is difficult because the trail of evidence is cold, and relating cause to effect becomes increasingly difficult. For practical policy purposes it is more important for evaluation to be timely than perfect. Consequently programmes are usually evaluated just before they are finally completed by which time most constituent projects have at least enough results to indicate the general direction of achievement[47].

The department has privatised most of its laboratories and now only directly manages the relatively small Weights and Measures Laboratory. It still owns the National Physical Laboratory (NPL) but has contracted its management out to a private sector company. Nevertheless the NPL carries out about 80% of it's work for the DTI, which is naturally

[47] For a more detailed description of the way DTI operate the ROAME system and carry out evaluations see John Barber's paper *Evaluation of support by the UK Department of Trade and Industry for the exploitation of science, technology and innovation by UK civil industry* available on the DTI's web site (http://www.dti.uk/tese).

concerned to assure the quality of the research involved. To this end the department's assessment unit (see above) has from time to time evaluated specific programmes and there have been studies of the impact of some NPL work on the national economy. More recently a new approach has been adopted by which a group of peer reviewers consisting of eminent scientists under the aegis of the Royal Society visit the laboratory three times a year. In advance of each occasion they consult with the DTI (who provide a secretariat) on their agenda. Afterwards they report on their findings to the DTI.

5.4.2 The Department of the Environment, Transport and the Regions (DETR)

The present department is the result of a merger in 1997 of the former Departments of the Environment and Transport, however, the broad remit of the combined department has not changed fundamentally. The DETR is responsible for the environment (including conserving and managing wildlife resources), transport, the problems of rural areas and the countryside, land and property use, the needs of local communities, sustaining economic development and social cohesion through effective regional action, housing and construction.

The Departmental Research Strategy is couched in very general terms and requires it to 'take account of Foresight and foresight-like activities'. It aims to 'monitor and evaluate the effects of existing policies' and 'ensure that new policies...are evidence-based', 'encourage innovation' and develop the research programme in an open and consultative manner fully consistent with the OST guidance – see above.

The 1998 Comprehensive Spending Review shifted the research focus somewhat towards some high priority policy areas such as a more integrated approach to transport policy, energy efficiency and delivering Kyoto commitments, promoting wildlife and biodiversity and enhancing economic growth in rural areas.

Scientific activity in DETR covers all these areas. It is driven essentially by policy needs and research is commissioned and managed by policy directorates rather than centrally by the Chief Scientist's Group. It ranges from basic to applied, though more is towards the latter end of the spectrum. For example there is much basic research in the global warming and air quality research programmes and a good deal of applied research in the transport programmes.

Almost all the research is contracted out. The two main research organisations that the department formerly controlled – the Transport Research Laboratory and the Building Research Establishment – have been privatised, but a substantial proportion of their work (most of it in the case of BRE) comes from the Department.

The ROAME system is nominally still in operation but, in practice, is applied very variably and often only in part. DETR argue that ROAME best suits a department where commissioned research is aimed at stimulating and supporting industry and a rationale for not leaving the work to the private sector is required. Where work is more directly supportive of the department's own policy no such rationale is needed because the activity arises directly from the department's own needs. Hence, DETR do not use ROAME to any extent at the approval/appraisal stage or as a true management tool.

ROAME statements are produced where policy divisions find it useful, but the residual requirement for them is not enforced in practice because in many cases they are seen as unnecessarily bureaucratic. The department, however, attaches importance to targeting

programmes and projects properly by means of well formulated objectives. Most monitoring, based on six-monthly reports, is routinely carried out by officials, who check progress against milestones and, if appropriate, make occasional site visits.

In some cases evaluation takes place at programme completion. However, research programmes usually continue over long periods (for example, there has been an air quality research programme for many years), so the rule is that there should be an evaluation at least once every five years. This generally looks at both outputs and impacts but, in practice, it is often difficult to attribute long term impacts to particular pieces of research. Such evaluation adopts a variety of methodologies, depending on circumstances. Usually the approach is similar to that applied in the DTI and includes an assessment of the extent to which objectives have been achieved. In appropriate cases it will include an element of peer review and, sometimes, bibliometric analysis.

There are also cases where policy and programme are evaluated – or 'reviewed' - together and the tendency towards this is increasing. In such cases the work may be done by one of the three in-house 'business consultancy units'. Apart from this the department has no internal programme evaluation capability but relies on external evaluation by independent experts. These are contracted directly to the managers of the programmes under scrutiny. To encourage independence an evaluation co-ordination officer within the Chief Scientist's Group has some involvement with the appointment of consultants and the assessment of conclusions.

The results of evaluations are circulated within the department often by distributing executive summaries to heads of units likely to be most interested. Whether sufficient resources are devoted to this aspect to ensure that more general lessons with across-the-board significance are always sufficiently highlighted is not entirely clear.

5.4.3 Ministry of Agriculture, Fisheries and Food (MAFF)

In summary MAFF's aims, stated in its PSA, are to ensure that the quality of food meets consumers' requirements; to encourage sustainable agriculture, fishing and food industries and a thriving rural economy; and to protect the public and the environment. For some of these aims MAFF shares responsibility with the DETR and the Food Standards Agency. In pursuit of its aims the ministry has ten objectives each supported by performance targets. All may involve scientific issues in relation, for example, to reducing the threat to health of animal diseases transmissible to humans, sustaining the rural environment (as indicated by reversing the long term decline in the number of farmland birds), conserving fish stocks while sustaining the sea fishing industry and reducing risks from flooding and coastal erosion.

In August 2000 MAFF published a 'consultation document' on its Research Strategy 2001-2005[48]. This sets out its Mission Statement for Science as to encourage and fund 'pertinent science to inform and implement its current policies, to provide scientific foresight, and to contribute to the identification of future policy options'. As indicated above the ministry sees comprehensive canvassing of scientific opinion as its first line of quality assurance. For this reason it has also set up a new Science Committee to advise it over the whole field. One third of the members of this committee, which is chaired by the ministry's Chief Scientist, are external independents.

[48] Ministry of Agriculture, Fisheries and Food (2000), *MAFF Research Strategy 2001 – 2005.*

The overall size of the programme is determined by the Comprehensive Spending Review as described earlier. The current annual expenditure on the R&D programme is about £105m. The budget is managed by the Chief Scientist's Group (CSG) on behalf of the Policy Divisions who act as customers for the research.

Most research is described as 'applied strategic' and about 1500 projects are in hand at any one time carried out in a number of programmes. Projects are usually of at least three year duration and sometimes last 10 to 12 years. A substantial amount of work is placed with MAFF's agencies - the Centre for Environment, Fisheries and Aquaculture Science; the Central Science Laboratory; and the Veterinary Laboratories Agency.

Research is managed within the ROAME framework. There is a more centralised approach than that in the DETR. MAFF operate a system known as the 'double lock' to ensure that all decisions involve both the relevant policy division and the CSG. This starts at the stage of appraising proposals when external peer review is also usually involved.

MAFF regularly reviews and evaluates areas of its research programme. These are ex-post assessments timed to take place at, or soon after, programme completion or, typically in the case of longer running programmes, every three years. Specific evaluations may be timed to coincide with critical points such as regulatory expiry dates or other special considerations. To ensure that analyses are independent, objective and critical, the majority of evaluations are carried out by external researchers selected by competitive tender. They involve MAFF policy customers, research contractors, independent experts and other stakeholders, such as industry or environmental groups. Reviews consider the quality of research undertaken and its value to MAFF policies and play a major part in directing future research in each area. They adopt similar methodologies to those employed in the DTI and the DETR.

5.4.4 Department of Health

The Department of Health (DOH) funds research under two main budgets. The larger – the NHS R&D Programme spends around £450m. per annum about half of which is used to support improvements in medicine and health under local control, mainly in NHS hospitals. About two thirds of the remainder - £150m - £200m – contributes generally to the medical science infrastructure. Some of this is applied research and some basic/strategic. The rest – about £70m – is managed by the department, centrally or through its regional offices, and used to identify and address particular research needs.

This latter part of the NHS R&D Programme is relevant to the current paper, but the smaller Policy Research Programme is more so. The DOH spends about £30m per annum on the latter, all of which is directly commissioned and mainly consists of applied research. The programme 'aims to help ensure that decision making in the policy domain is informed by all available and robust scientific evidence'. To this end it provides, 'through high quality research, a knowledge base for health services policy, social services policy and central policies directed at the health of the population as a whole'. The Programme extends across a range that includes 'population studies of health and social well being, lifestyle issues (related, for example, to smoking, sexual health and nutrition), promotion of health, prevention of illness, environmental factors (e.g. UV radiation and air pollution), social care for adults and for children, health service organisation and strategies for the health care of patients with particular diseases or conditions'.

The DOH has a number of scientific advisory committees, which advise on new scientific developments and present recommendations for further work. These now publish annual reports, minutes, agenda and conclusions. Some have also held open meetings and involved members of interest groups and consumer organisations. All projects commissioned externally are undertaken on the basis of detailed protocols which are peer reviewed.

Throughout the department's research programmes, policy directorates are increasingly making use of systematic reviews which bring together national and international research findings on particular topics.

The department has recently reviewed the ways in which it obtains scientific advice and has funded research to evaluate different methods of public participation in the advisory process. At all stages - before, during and after commissioning research - it relies primarily on widespread consultation and openness to ensure scientific quality. It invites international contacts to act as referees. It networks extensively and makes increasing use of the internet. In ways such as these and by participation in working groups and committees sponsored by international bodies the DOH draws on national and international experience.

At the appraisal stage the department has adopted a, horizon scanning, approach to identify issues for the NHS programme. This is systematically undertaken by a university based team of researchers. The Medical Research Council also provides independent advice on medical issues of concern. The department does not use the ROAME system, arguing that it is inappropriate to the structure of its research activities.

Monitoring arrangements during the lifetime of projects vary for the NHS programme. For the Policy Research Programme, officials maintain close contact with the research contractors during the course of a study.

The department does not institute formal *ex post* evaluations but concentrates on *ex ante* appraisal because they believe this focuses their activities more effectively. However, they encourage publication of research results in peer reviewed academic journals as a means of maintaining quality. Research projects are also reviewed by advisory committees. Once such peer review is completed successfully, researchers are encouraged to make their work widely available, using a number of dissemination routes, including a National Research Register (set up with NHS R&D funding) and the department's website.

The DOH is now also responsible for the Food Standards Agency (FSA), a relatively new body formed from an 'amputated limb' of MAFF in order to be demonstrably independent. It aims to protect public health from risks arising from food. Its current annual research budget is about £21m. It is still in the process of developing its procedures. The FSA is operating a programme-based ROAME system. Each programme represents a priority area and consists of between six and twelve research projects. The latter are all contracted out and selected by a combination of officials and external peers. Projects are monitored by officials on the basis of six monthly reports and occasional visits. Results are evaluated by external referees.

5.5 Findings and Conclusions

5.5.1 Prerequisites and impediments

Both those in central departments of UK government and those in the more 'executive' ones are clear that the quality of scientific knowledge for government purposes is critically dependent on being well focused and on 'asking the right questions'. They see it as essential to establish a culture where this occurs. Two elements of the UK public sector science environment are relevant here. On the positive side, the ROAME system with its emphasis on establishing clear objectives and on evaluation has been a focusing influence for almost 15 years. In a general, though not always specific, sense it is now ingrained.

On the negative side, UK government has a traditional tendency to secrecy. This has made it less easy to ensure that policy makers ask the right scientific or technical questions because it has inhibited the constructive participation of the wider scientific community. Identifying the key issues and questions, and selecting appropriate research teams, proposals and advisers are, themselves, tasks that cannot be left to policy makers alone, but require scientific input. So does the evaluation and interpretation of results. There are signs, however, that the culture of secrecy is beginning to break down as scientific and technical developments encourage more extensive involvement with the relevant professional community. Nevertheless, it is doubtful whether officials have yet fully absorbed the implications of this new environment, nor have the academic community yet sufficiently taken advantage of the opportunities for participation.

5.5.2 What works and what does not

In seeking to inform their activities by reference to the most relevant and the highest quality scientific knowledge that is available UK government departments have become increasingly eclectic with regard to the sources of such knowledge. Thus they acquire relevant information via research, specially commissioned for the purpose; they draw on research already in the public domain (e.g. in the literature); and they seek advice from appropriately qualified scientists, who base their input on pre-existing research undertaken by themselves or by others. Departments seek to recruit excellent scientists, for example, to advisory committees, but also, increasingly, to access the widest practicable range of scientific opinion.

Recognition of the latter need has led recently to more transparent processes, more extensive publication of government documents and wider consultation. This has been substantially facilitated and enabled by the advent of the internet. The process has still some way to go. For example, there is still a tendency to seek advice from national rather than international sources, but there are signs of change.

A parallel process has been evolving in relation to central guidance on the management of government science and technology. In the late 1980s and early 1990s there was some move towards increased intervention by the central departments but this has been found to be too inflexible. The present approach is generally to promote, centrally, some broad principles within a framework of strategic priorities, but to leave the detailed arrangements for individual departments to adapt to their particular needs.

Departments, both central and executive, believe that this loosely guided but mainly decentralised approach results in more appropriate procedures than does a more centralised system. It is true that it is, at least theoretically, more vulnerable to occasional laxity but in an increasingly critical climate this is less likely than it used to be. Recent large and well

publicised scientific crises – such as the BSE disaster – have themselves played a significant part in alerting government to the risks and in persuading them to adopt a more open approach in their own self-interest.

We have noted that the ROAME system, or something akin to it, is widely (though not universally) adopted and has had a generally favourable impact. Now, however, it is being applied in more flexible ways, each department adapting it to their particular needs and circumstances. In addition, it has been widely recognised that, while verifiable objectives are generally desirable, absolute insistence on this can lead to trivialisation.

Most departments now undertake regular *ex post* evaluations of programmes and policies. These tend to be carried out at, or soon after, programme completion both because results are required quickly and because later evaluation runs into problems of attribution and data collection. In the case of longer running programmes evaluations occur at regular intervals, usually of from three to five years. One department concentrates primarily on *ex ante* appraisal and subsequently peer reviews results as well as encouraging publication of research in peer reviewed academic journals as a means of maintaining quality.

Methodology is adapted to circumstances. In the case of programme evaluation a rough rule of thumb is often applied to the effect that the cost of assessment (including appraisal, monitoring and *ex post* evaluation) should not exceed 1% of programme cost. On the other hand programmes with, for example, high political profile or which have potential for replication elsewhere may get more attention.

In some areas *ex post* evaluation takes the form of regular set piece evaluations. These usually concentrate on assessing impact, often in relation to the achievement of previously stated objectives. They usually also include some assessment of the quality of science involved based mainly on direct peer review. In some case there is also bibliometric analysis and occasionally other quantitative indicators, such as patents, have been employed. No generally accepted system for measuring excellence in quantitative terms exists, however.

Most departments employ outside consultants or academics to carry out evaluation in order to establish independence. Where departments use in house evaluators they attempt to set them apart from the line management concerned for the same reason, though this may not always be completely effective. Such an arrangement, however, has advantages for data collection and cultural understanding.

In the case of research assessment the unit of evaluation has often been the programme but there is a tendency to see this as too limited. There are difficulties in applying conclusions across programmes and, increasingly, wider perspectives are seen as appropriate. Much importance is attached to evaluation feedback.

In many cases, however, evaluation is less formal, but more exclusively focused on research excellence. Here there is almost total reliance on peer review which is mediated through a plethora of advisory committees and a wide range of consultation processes. In this context the most notable development of recent years is that brought about by the internet. This has made it possible to tap into the scientific community at large to an unprecedented extent. Thus research opportunities can be widely publicised, sometimes with more than one iteration, for initial comment leading to better defined invitations to tender. Resultant proposals can be opened for further comment and then appraised and monitored by scientific

peers often on advisory committees. Finally, results are also subjected to both formal peer review and more informally tested against wider peer opinion over the internet.

Overall, the UK government's experience of seeking excellence in the scientific advice it receives is leading it less towards new techniques or methodologies than towards some important cultural changes. Some of these changes such as the development of a more focused and evaluation oriented approach have arisen partly out of the ROAME system and have been emerging for several years. Others, such as the moves towards greater transparency, openness and wider consultation of the scientific community are more recent developments and are the results, partly of public pressures, and partly of the increased complexity of the issues involved. These influences will continue to press. The system will continue to evolve.

5.6 List of interviewees

Dr. Philip Davies	Deputy Director Policy Evaluation, Cabinet Office
Mr. Stephen Morris	Principle Research Officer, Cabinet Office
Mr. Donald Franklin	Central Operations Research and Economic Division, HM. Treasury
Ms. Isabel Argimon	Central Operations Research and Economic Division, HM. Treasury
Dr Anthony Burne	Head of Research Policy and International Division, Ministry of Agriculture, Fisheries and Food
Dr.Ken McCowan	Research Policy and International Division, Ministry of Agriculture, Fisheries and Food
Dr. Susan Popple	Research Policy and International Division, Ministry of Agriculture, Fisheries and Food
Mr. Karl Cunion	Science and Technology Policy Division, Department of the Environment, Transport and the Regions
Mr. Alan Paterson	Science and Technology Policy Division, Department of the Environment, Transport and the Regions
Mr. John Barber	Director, Technology, Economics, Statistics and Evaluation, Department of Trade and Industry,
Mr. Stephen Pugh	Head of Research Coordination Unit, Food Standards Agency
Dr Chris Henshall,	Deputy Director of Research and Development, Department of Health
Mr. Edmund Quilty,	Head of Transdepartmental Science and Technology Group, Office of Science and Technology
Dr Andrew Wallard,	Deputy Director, National Physical Laboratory

5.7 Bibliography

Barber, J. (1999) *Evaluation of support by the UK Department of Trade and Industry for the exploitation of science, technology and innovation by UK civil industry.* Available at: http://www.dti.uk/tese

Cabinet Office (1989), *R & D Assessment: A Guide for Customers and Managers of Research and Development*.

Cabinet Office (1999) *Modernising Government,* Cmnd. 4310 Stationery Office.

Cabinet Office (1999) *Professional Policy Making for the Twenty First Century.*

Cabinet Office (2000) *Adding it up – Improving Analysis and Modelling in Central Government.*

Cabinet Office (2000), *Wiring it up: Whitehall's Management of Cross-cutting Policies and Services.*

HM Treasury (1997) *'The Green Book': Appraisal and Evaluation in Central Government.* Second Edition.

Ministry of Agriculture, Fisheries and Food (2000), *MAFF Research Strategy 2001 – 2005.*

Office of Science and Technology (1997), *The Use of Scientific Advice in Policy Making: A Note by the Chief Scientific Adviser.*

Office of Science and Technology (2000), *Guidelines 2000: Scientific Advice and Policy Making.*