Elliot Stern

What do we know about the utilization of evaluation? [1]

Based on the available literature on evaluation use[2]

Introduction

What follows might be better regarded as a literature mapping exercise rather than a literature review. However, sufficient material has been considered to begin to summarise what is known about evaluation use from available published sources.
We have drawn together material from three main sources:

- Academic and scholarly work in evaluation and related research.
- Professional guidance and advice on what evaluators and commissioners of evaluation should do.
- Examples drawn from the world of practice across different countries and institutions.

The sources reviewed fall into a number of broad categories, these are:

- Evaluation literature on the utilisation of evaluation.
- Policy studies about the institutionalisation of evaluation within governments and public administrations.
- Guidelines on evaluation practice and standards.
- Examples of specific evaluation arrangements that are relevant for use in different public administrations – national and international.

Literature on the utilisation of evaluation

There has debate about the use of evaluation for over twenty five years. This is primarily associated with two main scholars: Carol Weiss and Michael Quinn Patton. In the mid-1970's Carol Weiss began to focus on evaluation use, criticising simplistic notions of instrumental use. Based on empirical studies on how policy makers use evaluation, Weiss has been associated with a complex understanding of decision making and policy making in which evaluation findings are internalised, selectively used and rarely lead directly to specific

[1] Copyright reserved Elliot Stern 2003 and 2005: Can be cited if excerpts and quotations are acknowledged.
[2] Part of this review was supported by European Commission, DG Budget. Their support is gratefully acknowledged. All interpretations and conclusions are the responsibility of the author.

decisions or changes in policy. This theory, which is sometimes labelled ‘an enlightenment view of evaluation’, takes a long term incremental view about the way evaluation findings feed through to policy making. In this regard she is close to Peter Rossi’s understanding of the conceptual use of evaluation i.e. ‘the use of evaluations to influence thinking about issues in a general way’ (Rossi & Freeman 1985). She challenges the rational model of decision making and bases her conclusions on studies of how organisations actually work. She effectively argues for cumulative learning across many evaluations rather than direct use from particular evaluations. In some of her more recent work (Weiss, 1999) she associates herself with arguments of Sabatier and Jenkins-Smith (1993) about the importance of policy forums that bring together academics and policy makers and can act as a vehicle for the absorption of evaluation findings. She nonetheless remains true to her earlier arguments that evaluators should never expect that their inputs can override political agendas and administrative necessities which may push decisions in quite different directions from the recommendations of evaluators.

Michael Quinn Patton is more committed to the instrumental purposes of evaluation. ‘By utilization I mean intended use by intended users’ (Patton 1986, 1997). He places considerable emphasis on understanding the priorities of decision makers, engaging with them, encouraging them to own evaluations and their results in order to enhance use. Furthermore, Patton tends to emphasise the example of individual decision makers, particular people and not the decision itself. Indeed he is interested in individual ‘decision maker’s cognitive style and logic’. Unlike Weiss he is less interested in organisational and administrative processes and more with the act of decision and the particular decision maker.

It has been commonly observed (see for example Alkin 1990) that an important determinant of the difference in perspective between Weiss and Patton is their respective fields of practice and study. Patton has worked mainly in local community and voluntary organisations, and where he has worked at public administrations they have also been at a local level. Weiss has been more occupied with large-scale national programmes (beginning for example with the 1970’s War on Poverty in the US). This probably goes a long way to explaining the one’s preoccupation with complex organisational processes and the other’s preoccupation with individual decision makers and their decisions.

Although the above summarises, in a simplified way, the main debate between Weiss and Patton, both have contributed far more to our understanding of evaluation use. In particular Patton is probably the originator of the term ‘process use’ i.e. ‘changes…that occur among those involved in evaluation as a result of learning that occurs during the evaluation process’ (Patton 1997). Considerations of the process aspects of evaluation open up discussions also of different, more action orientated, views of evaluation purpose. For example, Patton

also discusses ‘intervention orientated evaluation’, ‘participatory evaluation ‘ and ‘empowerment evaluation’. These kinds of evaluation approaches are less relevant to the immediate concerns of this particular study, but do raise interesting methodological questions about how in the course of evaluations one can increase commitment and ownership by attending to the details of the evaluation process.

Weiss, especially in her early work, opens up another aspect of evaluation use, the importance of sound methodology and reliable data (Weiss & Bucuvalas 1980). Indeed in some of this early work Weiss anticipates some very contemporary discussions about evidence based policy making and the use of experimental methods. It should be acknowledged, however, that in later years as she investigated more carefully, the actual uses made of evaluation she shifted towards the position outlined above that we had to understand more about use in an organisational context. In her concern for sound methods and valid research, Weiss is more ‘hands on’ than many of her contemporaries thus Donald Campbell for example sees use for being essentially part of politics and of little concern for evaluators who should be driven by ‘truth’.

In many ways we can perceive a pendulum swing between organisational and implementation concerns on the one hand and methodological and quality assurance concerns on the other, back and forth over the last twenty five years. The current interest in evidence based policy, the rediscovery of ‘the gold standard’ of experimental methods and randomised control trials (RCTs) and the ‘realistic critique’ of such trials and experiments, represents a renewed movement toward the methodological end of the pendulum. Thus the ‘What Works’ school (see Davies, Nutley and Smith, eds. (2000)) is pre-eminently concerned with the nature of research evidence and the need for rigorous methodology. The adherents of this school often regard RCTs as the most reliable form of primary evidence and systematic reviews and meta-analysis – which make quantitative estimates of the overall effects of interventions by aggregating primary studies - as the most reliable approach for secondary evidence. However, already in Davies, Nutley and Smith (op cit) the editors of this collection themselves begin to address questions of better understanding the policy process. This is taken further by Huw Davies and Sandra Nutley (2002) when they move beyond data and findings to consider how evidence is disseminated and incorporated into practice. Thus these authors also follow the pendulum between methodological and implementation concerns.

However, the central components of the evidence based policy debate remain methodological. For the purposes of this literature mapping we would highlight, in particular, the debate around meta-analysis and systematic reviews (Gallo (1978), Pawson (2002)). The

core of this debate is the importance of drawing together evidence from previous related initiatives before embarking on new initiatives. For some, narrative reviews that are essentially structured qualitative literature reviews drawing together lessons, is their preferred approach. For others, they look to quantifying precise effects by looking to quantitative and experimental studies along the lines of RCTs in medical drug trials. Of course, the possibility of undertaking such quantitative studies is limited to a subset of policy areas usually in the social welfare, social benefit and welfare to work area. Despite the methodological disagreements among the proponents of different approaches to such secondary analysis, there is a widespread consensus that brings together evidence from across many evaluations is an important guarantor of the reliability of findings.

The most recent attempt to revisit the issue of evaluation use in a comprehensive fashion, is a publication of the American Evaluation Association – *New Directions for Evaluation* ('The Expanding Scope of Evaluation Use', Caracelli & Preskill, eds (2000)). The initial chapter attempts a reconceptualisation of evaluation use in terms of 'an integrated theory of influence'. This largely brings together much of the work referred to above distinguishing for example between process use and results use, intended and unintended use and a policy cycle time dimension, i.e. immediate end of cycle and long term use. Although this chapter brings together various previous attempts to address evaluation use within a single framework, it does not radically change the broad parameters of the debate outlined previously.

There are two additional chapters in this collections which should be referred to for the purposes of this mapping exercise. In Chapter 2, Preskill and Torres consider 'the Learning Dimension of Evaluation Use'. These authors highlight, in particular, learning at an organisational level that occurs as part of the evaluation process. They draw on 'constructivist' theories of learning, in which learners themselves engage in an active process of interpretation and dialogue in order to construct their own meanings. They link this perspective with Lave and Wenger's (1991) notions of 'communities of practice' i.e. groups of individuals working together, who are interdependent and whose tacit knowledge and problem solving capacities are integrated into their social and professional life. The authors suggest that learning from evaluations 'will most likely occur when evaluation is collaborative, is grounded in constructivist and transformational learning theories, and builds communities of evaluation practice.' The implications of this argument for evaluation use is to reinforce the importance of developing communities of both evaluators and users within organisations, if evaluation is to become part of an active learning process. It also steers us away from the narrow view of intended use that certainly informed the earlier work of Patton. Given that communities of practice will interpret and construct their own meanings using data and

findings that they bring into their own context, the use of evaluation may not be as evaluators or the commissioners of evaluation originally intended. However, transforming such evaluative outputs and processes into their own organisational context still constitutes evaluation use.

Policy studies on the institutionalisation of evaluation

There is an extensive literature around the role of evaluation in policy systems. This literature is largely based in policy analysis and public administration rather than in evaluation per se. The preoccupations of this body of literature include organisational learning in the public sector, how to build evaluation into public administrations, the nature of effective evaluation capacity, institutionalising and professionalising evaluation and analyses of the policy process itself. This latter set of preoccupations is the closest 'cousin' to the earlier discussion of evaluation use. However, within this literature the starting point is the nature of the policy process which then moves on to considerations of evaluation practice whereas within the earlier cited literature, the starting point is usually evaluation practice which then moves on to considerations of policy making.

Organisational learning has become a common metaphor within studies of various kinds of organisations, although most of this work has centred on the private sector. Leeuw, Rist & Sonnichsen, eds. (1994) bring together a range of experiences from the public sector across Canada, Sweden, the Netherlands, Norway and the United States that seeks to apply the concept of organisational learning through evaluation to public sector institutions. Among the themes highlighted in this study is the role of internal evaluations within public bodies (see especially chapters by Sonnichsen & by Mayne). The authors highlight the important role that can be performed by internal evaluation offices and how these can influence the overall organisations' evaluation practice and learning. They suggest that developing a 'double-loop learning' process that reports not only to programme managers but also to top management within government departments is an important part of the contribution that evaluation units can make[3]. However, these studies also suggest that whilst internal units can make an important contribution to organisational learning, they depend ultimately on what Sonnichsen calls 'a disposition towards critical self-examination'. His notion that 'self-

[3] It should be note that this use of 'double-loop learning' is unusual, even though this citation is true to the quoted authors. From Gregory Bateson, who originated many of these concepts in the 1950s and 1960s through to more contemporary authors who have come to be identified with them, such as Donald Schon and Peter Senge, 'double loop learning' is usually about learning to learn, by understanding higher order rules that can be applied across contexts.

reflection is crucial before organizations begin to learn' highlights the importance of creating a general evaluation culture within an organisation.

In the concluding chapter of this collection, Ray Rist identifies two sets of preconditions for learning from evaluation. The first set arises from the importance of fitting in with the policy cycle. This recognises that organisations need 'information at different phases of the policy cycle'. Synchronising evaluation outputs with different policy needs over time is seen as an important means of encouraging learning. The second set of preconditions for learning emphasises how information is transmitted and filtered within public organisations. Thus studies within this framework suggest that 'governmental organizations appear more receptive to information produced internally than that which comes from external sources'. This is especially so when the news is bad! Bad news is easier to receive from internal rather than external evaluators. Another precondition appears to be the credibility of the sources of information. This may on occasions favour internally generated evaluation findings, but may also depend on who are the sponsors or gatekeepers who bring information in to a public body. This can be seen as related to wider issues of relationship building and trust. The existence of such organisational attributes seems an important preconditions for receptivity to evaluation findings and to organisational learning within the Public Sector.

Another collection we have considered, focuses on evaluation capacity (Boyle & Lemaire, eds. 1999). Although some of the authors in this volume overlap with the Leeuw et all. collection referred to above, the concerns here are broader. Building evaluation capacity is seen in terms of 'national evaluation systems'. Thus evaluation capacity goes beyond the internal organisation of public bodies to include the location of evaluation in the executive or the legislature (e.g. Parliament) and broader issues of governance and institutional arrangements. For the purposes of this literature mapping exercise a number of themes explored in this collection are particularly relevant.

One such theme is the design of evaluation functions and offices. Richard Sonnichsen considers the advantages and disadvantages of a 'centralized' and a 'decentralized' model of evaluation functions. He note the potential for greater independence and credibility of centralized functions and their ability to develop strategic evaluation plans. This model has been favoured in Canada for example since the late 1970s. The 'downside' of centralized units and functions is also recognised. They are often seen as threatening by other units of administration with attendant resistance to change and potentially strained relationships.

Decentralised evaluation units are most likely to be intended to support decision making and programme effectiveness at a programme level. They are usually aligned with programme

management and are hence less threatening. On the other hand issues of independence and bias can arise from their closeness to programme personnel. Another weakness may be lack of methodological skills in evaluation. However, the most important criticism raised by Sonnichsen is the possible lack of power that decentralised units have, especially where decisions about programme and policy futures are still made at a centralised level. Ultimately this debate resolves itself into one of evaluation purpose. Where the purpose is primarily to improve programme and policy implementation there appear to be strong arguments for a decentralised model which will also favour learning at a decentralised level. Where the purpose is primarily to support central strategies and policy making the argument for a centralised model appears to be stronger – here the learning would tend to occur centrally rather than at programme or policy division level. Issues of professional competence and skills acquisition appear to be stronger within a decentralised framework. This issue is further discussed in the same volume by Richard Boyle who discusses the Human Resources aspect of evaluation specialists and the professionalisation of evaluation. He highlights in particular, the importance of appropriate training courses and education curricular, the way expertise is deployed mixing different skills together and the importance of continuous in-service training. From the point of view of this review much of this discussion is generic rather than use focused. However, from a use perspective Boyle makes a strong argument for 'developing evaluation users'. Evaluation users also need to be trained and developed to become 'consumers'. In some ways this can be seen as one of the consequences also of introducing managing for results approaches in public service organisations. This, in Boyle's terms creates the link between the supply and demand sides of evaluation use.

The shift in the institutional practice of evaluation within the public sector from monitoring and management to accountability and performance is widely noted. (See especially the chapter by Per Oyvind Bastoe: *Linking Evaluation with Strategic planning, Budgeting, Monitoring and Auditing*.) One definite tendency that this and other studies demonstrate is the integration of evaluation and monitoring with various approaches to performance management.

Among the elements of this approach (following from an influential OECD paper in 1995) are:

- objective and target setting,
- management responsibility to implement against targets,
- the monitoring of performance,
- the feed-in of such performance into future policy making and programming , and,
- the provision of information to external parliamentary and audit committees for ex post review.

There is also disagreement about the need and benefit of linking audit and evaluation, which Bastoe also recognises, especially when implementing performance management systems. It is, for example, important also to consider 'how learning actually takes place in organizations'. This attention to learning is also the pre-occupation of other authors in this field of the institutionalisation of evaluation in the public sector.

For example, a quite different approach to analysing evaluation use in policy settings is suggested by Peter Van der Knaap in his article 'Policy Evaluation and Learning: Feedback, Enlightenment or Argumentation?' (1995). He challenges the traditional rational – objectivist model of policy evaluation favouring rather a constructivist view in which policy makers conduct dialogues about evaluation findings in order to reach their conclusions. Thus 'policy-making is conceived of as an ongoing dialogue, in which both governmental and societal actors contest their views on policy issues by exchanging arguments'. At heart this argument challenges the 'positivist idea that policy evaluators may provide the policy-maker with neutral or objective feedback information or recommendations'. Rather than enlighten the policy maker 'At best, the evaluator might contribute to the quality of policy discourse by entering the negotiations that compose the policy-making processes with informed arguments and a willingness to listen, argue, and persuade or be persuaded'. This shift from the rational to the argumentative is, according to Van der Knaap, a way to 'institutionalise policy orientated learning'. This is not to suggest that the evaluator is relieved of the responsibility to provide reliable information and findings but needs also to supplement traditional analysis with material that will stimulate debate and allow different stakeholders to consider material presented from different perspectives. A similar logic informs a recent article by Ville Valovirta (2002). 'Rather than regarding evaluative information as indisputable knowledge, it is viewed as a collection of arguments, which can be debated, accepted and disputed'. According to the author utilisation of evaluation should be regarded as a process that runs through four stages:

1. Familiarisation with evaluation results and involvement in the evaluation process.
2. Interpretations based on expectations, assessments of the quality of research (truth test) and the feasibility of actions proposed or implied (utility test).
3. Argumentation in which 'individual interpretations are … subject to collective deliberation, discussion, negotiation and decision making'.
4. Effects which may take the form of decisions and actions, new shared comprehensions and increased awareness; and increased or undermined legitimacy.

Within this perspective on evaluation use:

'Evaluations force people to present well-grounded arguments for refuting evaluation conclusions and recommendations. This opens up possibilities for new understandings to emerge (Page 77)'

An important distinctions that Valovirta makes, is between the different contexts in which evaluation takes place. In particular he distinguishes between settings where there is a high level of consensus versus those with a high level of conflict; and settings where there is a low pressure for change versus those with a high pressure for change. He suggests that these contextual differences will determine the nature of the argumentation that takes place around evaluation findings.

Guidelines On Evaluation Practice And Standards

For many years there has been concern among evaluators about evaluation use. This has partly been fuelled by the academic debates reviewed above which have highlighted the problem of evaluations not being used. One approach to these concerns has been to produce guidelines usually for evaluators, but also for commissioners, that are intended to promote evaluation use. The earliest of these efforts have been produced by the American Evaluation Association and its precursor organisations, for example the Joint Committee on Standards for Educational Evaluation.

The Programme Evaluation Standards prepared as guidance to evaluators, particularly from a programme perspective, identifies standards under four main headings:

- Utility Standards
- Feasibility Standards
- Propriety Standards
- Accuracy Standards

For the purposes of this literature mapping exercise, utility standards which 'are intended to ensure that an evaluation will serve the information needs of intended users' are the most immediately relevant. These standards variously concern:

- being clear about stakeholders so that their needs can be addressed;
- ensuring the credibility of the evaluators so that their results are likely to be accepted;
- collecting relevant information from a broad range of sources as understood by clients and stakeholders;

- being clear about value judgement used to interpret findings;
- report clarity so that the information provided in reports is easily understood;
- reports that are well disseminated to intended users in a timely fashion; and,
- planning evaluations from the outset in a way that encourages follow through from stakeholders.

It is worth noting that the other standards also have implications for evaluation use. For example, Feasibility Standards include a sub-standard entitled *Political Viability*. This is concerned with obtaining the cooperation of different interest groups in order to limit bias or misapplication of results. Similarly the Propriety Standards include a sub-standard entitled *Service Orientation* which is designed to assist organisations to address and effectively serve the needs of a full range of targeted participants; and one entitled *Conflicts of Interest* which seeks to avoid compromising evaluations and their results. Accuracy Standards are also infused by notions of use, for example *Valid Information*, one of the sub-standards is justified in terms of assuring 'the interpretation arrived at is valid for the intended use'.

This set of standards has been widely imitated and adapted to different national contexts including, most recently, by the Deutsche Gesellschaft für Evaluation – DeGEval (German Evaluation Society). There are also discussions and plans in France and the UK to develop standards. These discussions have however moved away from directly adopting the North American model.

Alongside the widespread adoption of the AEA Programme Evaluation Guidelines there has also been the emergence more recently of a set of guidelines for the Ethical Conduct of evaluations. These differ from programme guidelines, being more concerned with the ethical dilemmas that both commissioners and evaluators face in the course of the evaluation process. Such guidelines have been variously prepared by the Australasian Evaluation Society (AES), The Canadian Evaluation Society (CES) as well as the AEA. These Ethical Guidelines are relevant to a discussion about evaluation use because they too are concerned with the credibility of evaluations and also the feasibility of the evaluation task.

Both the CES and the AEA direct their attention at evaluators. Thus the CES guidelines under the three main headings of *Competence*, *Integrity* and *Accountability* are concerned that *evaluators* should be competent, act with integrity and be accountable. Similarly the AEA - under its various headings of Systematic Inquiry, Competence, Integrity/Honesty, Respect for People, Responsibilities for General and Public Welfare and Recommendation for Continued Work – also direct their attention to what evaluators ought to do.

The set of Ethical Guidelines which stand out from the others already referred to, are those produced by the Australasian Evaluation Society (AES). These are concerned with ethical

behaviour and decision making among commissioners, users and teachers of evaluation as well as evaluators themselves. According to the AES the primary groups addressed by their guidelines are commissioners and evaluation teams or evaluators. In this regard also the AES diverges from some of the other guidelines referred to above. They acknowledge that evaluators often work in teams rather than mainly as individuals.

The AES *Guidelines for the Ethical Conduct of Evaluations* follows the evaluation cycle by grouping its guidance under three main headings:

- Commissioning anpreparing for an evaluation.
- Conducting an evaluation
- Reporting the evaluation results.

One view one can take about the AES guidelines is that they constitute a quality assurance framework for the entire evaluation process. These guidelines (which can be found at http://www.aes.asn.au/ethics_guidelines-1.pdf) are relevant to evaluation use because of the focus that they bring to:

- Questions of credibility;
- Shared expectations between evaluators and commissioners about what can be delivered through an evaluation;
- Strengthening the basis for evaluation judgements;
- Reducing conflicts during the course of the evaluation process;
- Ensuring balance and simplicity in the way reports are presented.

However it should be noted that the AES regards these guidelines as complementary to rather than as a substitute for other guidelines such as the Programme Evaluation Guidelines – indeed they encourage the use of the Guidelines for the Ethical Conduct for Evaluations jointly with the Programme Evaluation Standards.

There is a variety of other guidelines for evaluation that are relevant to this literature mapping exercise. For example, the Means Collection Volume 1 (Evaluation Design and Management) has a section on optimising the use of evaluation. This focuses on dissemination, distinguishing between both different communication channels (e.g. reports, synthesis, article, confidential note) and audiences (e.g. commissioners of the evaluation, steering groups, managers, European institutions, citizens and journalists). The Means Collection also briefly touches on different forms of use including for example:

- Raising levels of knowledge;
- Deciding on improvements; and
- Judging success and failiure.

The Means Collection authors also acknowledge the 'absence of a direct short-term link between recommendations and decisions'.

On the boundaries of the territory to be covered by this mapping exercise would be a wider literature on innovation and good practice. Thus in many studies that have been conducted at a European level there is a concern for dissemination of innovation, an interesting example of this is found in Simo Mannila et al, eds (2001). In this volume, Philip Sonderberg discusses 'Mainstreaming, Innovation and Knowledge Networks', the way in which innovations are spread is close to the problem of how to encourage the take-up of the results of evaluations. Sonderberg also focuses on knowledge networks (cf. Policy Communities and Communities of Practice) as a means of ensuring the dissemination and mainstreaming of new ideas. Arguably this broader literature on the diffusion of innovation could be regarded more generally as a source of insights into the use of evaluation. Suffice it to say for the purposes of this review, much of this literature is consistent in its thinking with the literature on evaluation use outlined in this mapping exercise.

Examples Of Specific Evaluation Arrangements

In this section a number of specific high-profile examples of evaluation use in public administrations are identified. By way of introduction it should be noted that these examples are not evaluated here and as far as we are aware have not themselves been evaluated by others. Most of the publicly available source material is either uncritical (and sometimes verging on the propagandist), or is aspirational because the initiatives concerned are themselves at a very early stage. It is fair to say that none of these cases are without problems, and these problems are acknowledged below where information is in the public domain.

The World Bank has developed an elaborate evaluation process that seeks to link planning and resource allocation to results. The evaluation process relies heavily on a rating system, using criteria such as relevance, efficiency, effectiveness, sustainability, institutional development and the Bank and borrowers performance. This system involves a number of different methods which include the Bank's own corporate scorecard and ICR (Implementation Completion Report). This is filled in by Programme Managers and independently monitored by the Bank's Evaluation Department. This evaluation process is embedded in a results based management framework, and indeed a broader risk assessment and control environment.

In order to foster a learning culture, the Bank has initiated a knowledge management strategy based on networks intended to draw lessons across countries and regions and

ensure that 'global best practices are used to shape country development objectives' (1998). Knowledge networks are made up of a number of different components including:

- Forming professional communities of practice;
- Developing an online management system ;
- Establishing help desks or advisory services;
- Making key statistics available online;
- Providing a directory of expertise;
- Tracking engagement or transactional information;
- Creating online dialogue spaces;
- Increasing external access to Bank knowledge.

The implementation of this system has been uneven, among the reasons for this have been the work pressure on staff and pre-existing 'information hoarding cultures' which have militated against the sharing of information across sectors and departments.

Similar difficulties have been found in linking planning and resource allocation to results. This has been somewhat improved by the gradual refinement of a results based management system within the bank. Thus in the latest Annual Report on Operations Evaluation (2000-2001), progress with knowledge management activities have been noted in a number of areas, in particular:

- The active role of the helpdesk which since 1999 has fielded 3000 enquiries about evaluation methodology, findings and recommendations. An interesting feature of this helpdesk is that it is accessed both by OED internal staff and external partners.
- Participation in evaluation conferences, sector weeks and knowledge fairs in order to diffuse evaluation findings and lessons as widely as possible.
- Collaboration across the Bank's different evaluation units in order to create 'an evaluation community of practice'. This provides a forum to exchange evaluation news, activity and products and there is now a shared web portal that collects together material developed by different evaluation units.
- Synthesising evaluation findings providing the basis for a 'Lessons Paper Series' examples of recent titles include, Recurrent Lessons in India, Utilization of Project Implementation Units, and the World Bank's Experience with Institutional Development.

One innovation in the Bank's evaluation practice which is of particular relevance in the context of this mapping project is the attempt to trace the influence of three of the past OED evaluations:

'the study reconstructs the circumstances under which experiences of use occur, what change they bring about, and what factors favor or impede the perception of usefulness associated with OED studies' (World Bank 2001).

These studies, which have not yet reported, were chosen to include cases where there was already some indication of utilisation as this was likely to generate greater understanding of the circumstances under which evaluation occurred.

The Netherlands has recently introduced a policy evaluation system that explicitly links budgets to operations and results. The new regulation 'Central government achievement data and evaluation research' came into force in all Dutch government departments in January 2002. This regulation overseen by the Ministry of Finance is the result of a interdepartmental working group which had two aims:

'to ensure that the evaluation function within central government is sufficiently guaranteed, and;
to ensure that the policy information supplied within the framework of the deparmental budget and the deparmental annual report complies with the quality requirements applying to this information.

In order to achieve these targets, the regulation contains further conditions regarding:

I. *the integrated deployment of evaluation instruments;*
II. *the weighing up that needs to be made when making use of ex ante evaluation research;*
III. *the degree to which policy is covered by periodical evaluation research and its frequency (completeness and periodicity);*
IV. *the methodological-technical quality of evaluation instruments and the way in which policy information is obtained;*
V. *the way in which civil servants and top politicians are informed about the results of periodical evaluation research, and;*
VI. *the responsibilities within a department for proper implementation of the regulation.*

(Policy Evaluation in the Netherlands: Linking budgets to operations and results, Peter van der Knaap 2001))

The use of evaluation is a feature of much of the system that is now being introduced in the Netherlands. This includes:

- Gathering together information on what has been achieved and the effects of policies.
- Ex ante evaluations which compare two or more policy options systematically.
- Ex post evaluations regarding 'achieving goals, effectiveness and appropriateness'.

With particular regard to the use of evaluation results, the new regulation requires that civil servants and politicians are informed about results. However, departments are not obliged to follow a standard process, but rather to specify what their procedure is 'in harmony with the departmental cycles of budgeting, planning and control'. In the Netherlands, the debate around evaluation, budgeting and performance management are also rooted in questions of legitimacy. As Van der Knaap argues this integration of evaluation and performance management: 'provides society with a clear answer to the question: what can we expect of the government. And this in turn helps to answer the question: how much are we willing to pay for this?'.

The Canadian Government has a long history of results based management and the incorporation of evaluation within that governments management framework (see for example chapter by John Maine in Leeuw et al. 2000 discussed above). The latest statement of this policy was published in April 2001 and replaced an earlier version published in July 1994.

'The policy is based on three fundamental principles:

- *first, that achieving and accurately reporting on results is a primary responsibility of public service managers;*
- *second, that rigorous and objective evaluation is an important tool in helping managers to manage for results; and*
- *third, that departments with the support of the Treasury Board Secretariat are responsible to ensure that the rigour and discipline of evaluation are sufficiently deployed within their jurisdictions.'*

Treasury Board Manual (2001)

Within this policy framework, evaluation has two main purposes;'

- *to help managers design or improve the design of policies, programs and initiatives; and*
- *to provide, where appropriate, periodic assessments of policy or program effectiveness, of impacts both intended and unintended, and of alternative ways of achieving expected results.'*

The whole emphasis of this policy is centred around use, evaluation is seen as a support for managers and decision makers. The Canadian government sees itself as 'committed to becoming a learning organisation' and the Treasury Board Secretariat which oversees government activities is obliged to:

'support the practice of evaluation in departments by providing advice on best practices, setting standards, monitoring the evaluation capacity in departments and using the products of evaluation to inform decision-making at the centre'.

These policies are integrated into the job responsibilities of administrative managers within government departments thus deputy heads are expected to provide 'early warning of evaluation findings' that indicate major concerns. Departmental heads of evaluation communicate their findings that raise management concerns to senior management. Whilst departmental managers are obliged to 'incorporate approved evaluation findings and measures for improvement into priority setting, planning, reporting and decision-making processes'. Furthermore, this policy is reinforced by evaluation standards that have been adopted for the Government of Canada. These include standards of evaluation reports that are intended to ensure that senior managers and external readers can understand the issues, contains clear and actionable recommendations and an analysis of the exposure to risks associated with these recommendations.

The current government in the UK has also made considerable efforts to improve the policy making processes of central government, in particular the Cabinet Office has put forward a framework for 'Professional Policy Making for the Twenty First Century' (1999). This framework is concerned with 'delivering outcomes' and to that extent shares many assumptions with a results based management approach. A key part of this approach is the use of evidence which follows from a commitment to concentrate on what works, a slogan used by the Government has been 'what counts is what works'. The commitment to evidence based policy involves improving access to research based findings through many of the methods which have been referred to elsewhere in this review. For example, knowledge pools and 'hubs' are seen as a means of knowledge management which provides access not only to results of research but also to 'evidence held in the minds of both front-line staff in departments, agencies and local authorities and those to whom the policy is directed'. This approach continues to inform many of the initiative of the UK government including wide ranging reform of the Civil Service which is intended to 'deliver results of high quality and good value'.

Any system that seeks to encourage the use of evaluation results, process use of evaluations and related practices of information management, organisational learning and

results based management will encounter problems. All the examples included in this section have such problems, but not because their conception is flawed. Rather it is because as the earlier parts of this mapping exercise have suggested, evaluation use is complex and embedded in organisational, policy, administrative and methodological realities. There is no quick fix, magic bullet or best practice solution that can be found and copied. This does not mean that there is nothing to learn from these cases, although the similarity in many regards to existing Commission practice can be seen as an initial positive lesson.

Conclusions

Evaluation use has been a topic of concern within the evaluation and policy analasis communities for over 25 years. This overview of the available literature has therefore highlighted long standing as well as more recent debates. What we have uncovered also fits well with other parts of this study: many existing aspects of Commission practice – in its recognition of instrumental, process and more subtle 'argumentation' uses, in its attempts to link evaluation with decision making, in its approach to the organisation and institutionalisation of evaluation within and across DGs – accord with what we know are important determinants of successful evaluation use.

We can draw together some of the main relevant conclusions without being too sensitive to what are often hard fought debates, mainly because the protagonists in these debates often themselves recognise that their differences arise from differences in context and fields of interest. The main conclusions we would draw are as follows:

- Evaluation use understood as the direct 'instrumental' use or take-up of results, findings, conclusions and recommendations is unusual although it does occur. The likelihood of such direct take-up will be shaped by administrative and policy imperatives as well as the quality of evaluation reports.
- Evaluation use is more likely to be cumulative across evaluations or in the case of particular evaluations be transformed by argumentation: the process of challenge and counter challenge to evaluation evidence and interpretation. Evaluation is likely to be influential rather than the sole cause of subsequent action.
- Even where results are not used or only used indirectly, the 'process' of evaluation initiation and management with all their attendant opportunities for discussion and reflection can be useful by offering opportunities to exchange information, clarify thinking and develop frameworks. These indirect uses of evaluation require that

evaluators provide inputs and arguments that evaluation users can feed into their debates and argumentations.

- Whatever the kinds of evaluation use are envisaged or occur, involving stakeholders alongside evaluators and administrators throughout the evaluation cycle is essential in order to build commitment to evaluation use.
- Sound methods are important. These should be suited to the evaluation object, context, purpose and questions. There are cycles in the advocacy of particular methods: one must conclude that in their place most are relevant. The contemporary interest in systematic reviews and the synthesis of evaluation findings is consistent with the consensus that most evaluation use occurs across several rather than any one single evaluation.
- To use evaluation findings and processes, administrations need to learn. In order to do so they need to build up their human resources and competencies in evaluation, create networks and 'communities of practice', institutionalise appropriately (sometimes in a centralised and sometimes in de-centralised ways - each have their strengths and weaknesses) and build up their knowledge management systems. A culture of evaluation, interpreted most commonly as a reflex of self critical openness, is widely seen as essential to organisational learning.
- The uses of evaluation are inextricably bound up with the different purposes of evaluation. Those who link evaluation closely to accountability, for example through results based management, take a different view of evaluation use than those who prioritise learning or knowledge production evaluation purposes. It could be concluded from some of the literature that these different purposes and their associated uses are difficult to combine in one system. This may be because of the contrary requirements of self criticism and openness required for learning and a more purposive and 'error-free' culture that is often associated with performance management.
- Standards appear to be one way in which quality is assured and use is made more likely. Many existing sets of standards in evaluation appear to be particular in their content and suited also to particular contexts. The most useful standards would address the needs of evaluators, administrations, and wider publics; would include technical and ethical components; and be integrated within the institutional logics where they were to be applied. Such standards would probably also include key principles that are known to encourage the diffusion of innovations.

- The brief review of actual cases of use strategies, suggests that many if not most administrations are facing some of the same problems of evaluation use as the European Commission and other European Institutions. However none appear to have solved the problem. It would appear that all the 'solutions' identified in this section need to be managed, integrated and institutionalised and will need constant re-engineering.

References and Sources

Alkin, Marvin C. (1990) *Debates on Evaluation*. SAGE.

Barbier, Jean-Claude and Bernard Simonin (1997) 'European Social Programmes : Can Evaluation of Implementation Increase the Appropriateness of Findings ?' *Evaluation Vol 3, No. 4, pp391-407*. SAGE.

Bemelmans-Videc, Marie-Louise, Ray C. Rist and Evert Vedung, eds (1998) *Carrots, Sticks & Sermons*. Transaction Publishers.

Boyle, Richard and Donald Lemaire, eds (1999) *Building Effective Evaluation Capacity : Lessons from Practice*. Transaction Publishers.

Canadian Treasury Board (2001) *Treasury Board Manual 'Review' Volume*. Canadian Treasury Board.

Caracelli, Valerie J. and Hallie Preskill, eds (2000) *The Expanding Scope of Evaluation Use*. Jossey-Bass.

Davies, H. and S. Nutley (2002) *Discussion Paper 2 Evidence-Based Policy and Practice: Moving from Rhetoric to Reality*. University of St. Andrews Research Unit for Research Utilisation.

Davies, Huw T.O., Sandra M. Nutley and Peter C. Smith, eds (2000) *What Works?*. The Policy Press.

Department of Transport, Local Government and the Regions (2000) *Evaluation Notes*. UK DTLR.

Deutsche Gesellschaft fűr Evaluation e.V. (2002) *Standards fűr Evaluation*. Die Deutsche Bibliothek.

Duffy, M. (2000) *HEBS Commissioned Research Reports: Style Guidelines*. HEBS.

Efficiency & Effectiveness Group (1999) 'Learning Lessons: The Role of Policy and Programme Evaluation' *Modernising Government White Paper (UK)* . Efficiency & Effectiveness Group.

Efficiency & Effectiveness Group (1999) *Modernising Government White Paper, Learning Lessons : The Role of Policy and Programme Evaluation*. EEG.

European Commission – EuropeAid (2001) *Manual Project Cycle Management*. European Commission.

European Commission Budgetry Overview and Evaluation Unit (1997) *Evaluating EU Expenditure Programmes*. European Commission.

Forss, Kim, Claus C. Rebien and Jerker Carlsson (2002) 'Process Use of Evaluations: Types of Use that Precede Lessons Learned and Feedback' *Evaluation Vol 8, No. 1, pp29-45*. SAGE.

Fulbright-Anderson, Karen, Anne C. Kubish and James P. Connell, eds (1998) *New Approaches to Evaluating Community Initiatives Volume 2, Theory, Measurement and Analysis*. The Aspen Institute.

Gallo, P.S (1978) 'Meta-analysis – a Mixed Metaphor' *American Psychologist 33, pp515-517*.

Gautam, Madhur (2002) *2000-2001 Annual Report on Operations Evaluation*. World Bank Operations Evaluation Department.

Guy, K., J. Clark, K.Balazs, J. Stroyan and E. Arnold (1998) *Strategic Options for the Evaluation of the R&D Programmes of the European Uniton*. Technopolis Group.

Hedges, A (1994) *Commissioning Social Research: A Good Practice Guide*. Social Research Association.

Horton, D., R. Mackay, A. Andersen and L. Dupleich (2000) *Evaluating Capacity Development in Planning, Monitoring and Evaluation: A Case From Agricultural Research*. ISNAR.

IST Monitoring Panel (2000) *Report of the 2000 IST Monitoring Panel*. DG Information Society.

Lave, J. and Wenger, E. (1991) *Situated Learning : Legitimate Peripheral Participation*. Cambridge University Press.

Leeuw, Frans L., Ray C. Rist and Richard C. Sonnichsen, eds (1994) *Can Governments Learn?*. Transaction Publishers.

Macintyre, Sally, Iain Chalmers, Richard Horton and Richard Smith (2001) 'Using Evidence to Inform Health Policy : Case Study' *BMJ Vol 322, pp 222-322*. BMJ.

Mannila, S., M. Ala-Kauhaluoma and S. Valjakka, eds (2001) *Good Practice in Finding Good Practice: International Workshop Evaluation*. Rehabilitation Foundation.

Means Collection (1999) *Evaluating Socio-economic Programmes*. European Commission.

Monitoring and Evaluation Department of MSF (2000) *The Utilisation of Evaluation Recommendations in Medecins sans Frontieres – Holland: A Study of 10 Evaluation Reports 1997-1999, Evaluation Series No. 14*. MSF.

Norwegian Ministry of Foreign Affairs (1999) *Evaluation Activities in the Norwegian Ministry of Foreign Affairs, Annual Report 1998*. Norwegian Ministry of Foreign Affairs.

Norwegian Ministry of Foreign Affairs (2000) *Evaluation Activities of the Norwegian Ministry of Foreign Affairs, Annual Report 1999*. Norwegian Ministry of Foreign Affairs.

Norwegian Ministry of Foreign Affairs (2001) *Evaluation Activities of the Norwegian Ministry of Foreign Affairs, Annual Report 2000*. Norwegian Ministry of Foreign Affairs.

Operations Evaluation Department (1999) *1998 Annual Report on Operations Evaluation*. World Bank Operations Evaluation Department.

Patton, Michael Quinn (1997) *Utilization-Focused Evaluation: The New Century Text, Edition 3*. SAGE.

Pawson, R. (2002) 'Evidence Based Policy : In Search of a Method *Evaluation Vol 8, No. 2, pp157-181*.

Pollitt, Prof. Christopher and Prof. Helen O'Neill (1999) *An Evaluation of the Process of Evaluation of EC External Aid Programmes*. COTA.

Preskill, Hallie and Rosalie T. Torres (2000) 'The Learning Dimension of Evaluation Use' *New Directions for Evaluation No.88*. Jossey-Bass.

Strategic Policy Making Team, Cabinet Office (1999) *Professional Policy Making for the Twenty First Century*. UK Cabinet Office.

The Research Council of Norway, Strategic Planning (1996) *Evaluation Strategy*. The Research Council of Norway.

UK Government (1999) *Modernising Government White Paper*. TSO.

Valovirta, Ville (2002) 'Evaluation Utilization as Argumentation' *Evaluation Vol 8, No. 1, pp60-80*. SAGE.

Van der Knapp, P. (2001) *Policy Evaluation in the Netherlands: Linking Budgets to Operations and Results*. Netherlands Ministry of Finance.

Van der Knapp, P. (2000) ‘Performance Management and Policy Evaluation in the Netherlands : Towards an Integrated Approach ‘ *Evaluation Vol 6, No. 3 pp335-350*. SAGE.

Van der Knapp, P. (1995) ‘Policy Evaluation and Learning : Feedback, Enlightenment or Argumentation’ *Evaluation Vol 1, No. 2 pp189-216*. SAGE.

Weiss, Carol Hirschon (1999) ‘The Interface between Evaluation and Public Policy’ *Evaluation Vol 5, No. 4, pp468-486*. SAGE.

Elliot Stern

---

ekonomista, politolog, antropolog, badacz - wybitna osobowość w dziedzinie ewaluacji programów publicznych. Członek zarządu w Brytyjskiej Akademii Nauk Społecznych, prezes m.in. Brytyjskiego Towarzystwa Ewaluacyjnego (1994-1996), Europejskiego Towarzystwa Ewaluacyjnego (2002-2003), obecnie prezes Międzynarodowej Organizacji do Współpracy w Ewaluacji (IOCE). Wykładowca akademicki, redaktor czasopisma Ewaluacja wydawanego przez Tavistock Institute, którego przez wiele lat był dyrektorem. Prowadził ewaluacje m.in. dla brytyjskiego rządu i w Brytyjskim Towarzystwie Ochrony Zdrowia. Ponadto jest także niezależnym konsultantem ds. ewaluacji przy takich instytucjach jak Komisja Europejska, Bank Światowy, UNESCO, Międzynarodowa Organizacja do Spraw Badań nad Rolnictwem (CGIAR).