The Agora and the Role of Research Evaluation

Frederiksen, Lars Frode; Hansson, Finn; Wenneberg, Søren Barlebo

The Agora and the Role of Research Evaluation

by

Lars Frode Frederiksen
Finn Hansson
Søren Barlebo Wenneberg

all from

Department of Management, Politics and Philosophy
Copenhagen Business School
Blaagaardsgade 23 B
2200 Kbh. N.
Copenhagen
Denmark

Abstract

Research evaluation is an activity undergoing change. The traditional peer review system with its focus on scientific content and methodology has long been the backbone of research evaluation, but over the last three decades other criteria and considerations have also been integrated into the evaluation of science. This paper investigates how recent societal developments — epitomised by the concept of the Agora — influence research evaluations and how we perceive them. Are they still grounded in a scientific rationale, or are they more to be understood as a result of a social rationale?

Keywords: research evaluation, agora, peer review, evaluation criteria, contextualisation, social and economic issues.

1. Introduction

The role of science in society is changing rapidly and is constantly discussed. Numerous social science theories have taken up the discussion, offering a variety of explanations for the change. Despite the many different explanations, it seems to be commonly accepted in the discussion of modernity that a change is taking place. However, how deep or how radical this change is, remains open to debate.

Are we witnessing the birth of a whole new concept of what used to be science (the 'legend of science' to quote Ziman 1994) and scientific knowledge? Is it a question of changes in the role or function of science in society such as a much more politically powerful and changed public understanding of science (Wynne 1996, Expert group (EU) 2002) or an erosion of the role and authority of traditional expertise (Beck, Scott & Giddens 1994)?

Alternatively, is it a question of a changed policy — of a changed role of science policy emanating from governments and influential international organisations (OECD, EU) towards a new role for scientific knowledge in the knowledge economy? In the last twenty years, the development in science policy shows very clearly that the political agenda has changed. Among other changes numerous new types of research evaluation procedures have emerged — all as the result of local, national or international political attempts to steer, regulate or redirect organisations, research fields or research agendas. This development can only be interpreted as a clear sign of a more powerful social and political interest, not only in distributing funds to scientific research but also in influencing what goes on in the scientific system. This is part of a more general development in all parts of the public sector, where visibility of activities and discussions with users is becoming a new organisational reality.

In any case, a large change in the relationship between science and society is taking place today as witnessed by a number of recent publications from the sociology of science field. The classic understanding of the purpose of science as a production of knowledge for its own sake is still alive, but new understandings of the application of science are under rapid development. Concepts such as 'mode 2 knowledge production' (Gibbons et al. 1994), the 'Triple Helix' (Etzkowitz & Leydesdorff 1997, 2000) and the 'entrepreneurial university' (Clark 1998) all point in the same direction: science is no longer the only important producer of certified or reliable knowledge and the application of scientific knowledge is not a question to be dealt with only by scientists.

'Re-thinking science' by Nowotny et al. (2001) is an important and the most recent contribution to the opening of the debate on the changed social functions of science in society as well as the internal consequences for science as a knowledge production system. The central concept in this book is the 'Agora' — a concept that conceptualises the relationship between science and society and does this in a new and radical fashion — and therefore also offers a new perspective on the evaluation of research.

2. Research question

The purpose of this paper is to discuss and develop the concept of the Agora introduced by Nowotny et al. (2001) to describe how society can produce socially robust knowledge in a situation where traditional methods of constructing scientific reliability are insufficient. We believe the concept of the Agora to be the most radical formulation of the ongoing changes in the relations between science and society, but until now most of the discussion of this new concept has been on a rela-

tively general level without actually addressing the more specific aspects of the role of the classic scientific evaluation system. With this paper we will open a much more detailed and practical discussion of the Agora, based on a thorough analysis of the relationship between the changing role of science in society and research evaluation as a means of serving various purposes in authorising scientific knowledge claims.

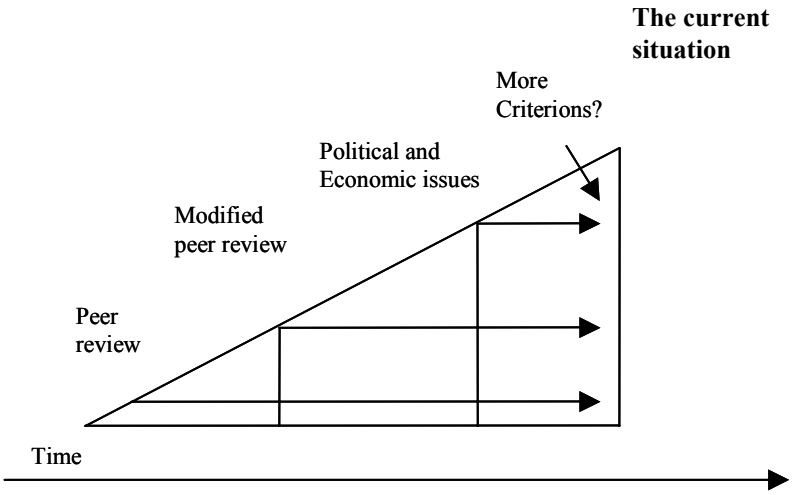
In order to be able to focus our discussion on the complex relationships between science in society and the different systems of evaluating the quality of scientific knowledge, we will refrain from a more detailed theoretical discussion of the role and function of the public sphere in relation to science. The public sphere and mediation through the media play an increasingly important role for an understanding of the ongoing changes between science and society. However**,** in order to maintain the focus on our central research question — the changes in the relationship between science in society and evaluation — the role and function of the public sphere and the public understanding of science (Wynne 1996) will be treated at a descriptive level only.

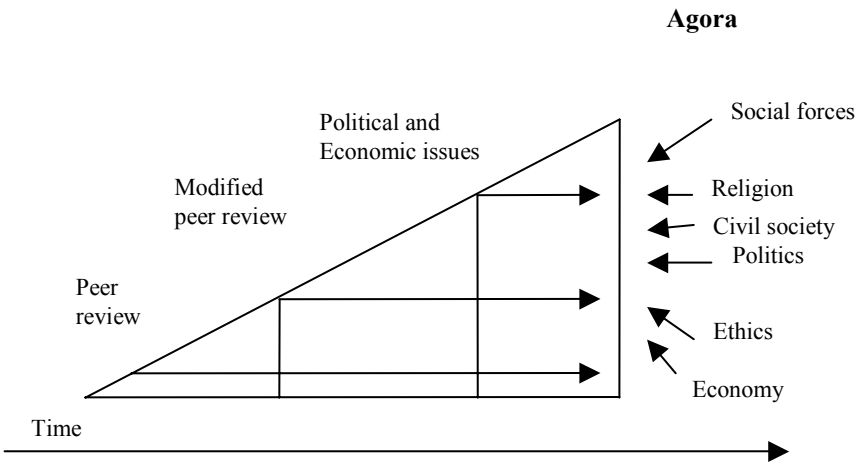
Our framework in this paper is first to elucidate the meaning of the Agora as a conceptualisation of an open marketplace or network with many different actors, political and economic interests and competing scientific knowledge claims. In order to clarify and exemplify the discussion of the consequences of the ongoing important changes in the relationship between science and society, we will discuss below the recent developments in the field of the evaluation of science and give a broad overview of this area. We will examine the traditionally recognised quality control system in order to establish an empirical background for understanding the changes in the relationship between science and society. Following on this broad discussion, we will relate the development of research evaluation as it appears in various specific forms with the recent change in the relationship between science and society. With the help of four selected cases, we will explain what we see as the most important systematic changes in the complex system of research evaluation and use the cases to describe how evaluations are changing in dimensions and functions. We will then argue that the four cases seen together constitute a continuum of complexity in how society today handles scientific knowledge claims — from the traditional peer review to the most complex benchmarking evaluation, a continuum of complexity that illustrates well the development of a new relationship between science and society.

In our view, the development can be understood in two distinct and very different ways. One is a continuation of an ongoing process that constantly increases the number of approaches to conducting research evaluation — giving more complex types of evaluations involving ever more criteria. The other possible understanding is that the appearance of the Agora means a real and radical change in the relationship between science and society and therefore also a change in the status of all the various approaches and elements of research evaluation. The figure below illustrates this difference. In our conclusion we will discuss which of these understandings capture the empirical situations most accurately.

In the text we will discuss the stages in the figure from the classical peer review to the current new complex evaluations. Starting with the classical peer review, we examine in turn evaluation as a tool for organisational learning and development, evaluation as a tool for measuring impact and thus creating input for politicians and other decision makers, and finally a suggestion concerning how evaluation may occur **today — either** conceptualised as a continuation of the earlier forms or conceptualised as a totally new **situation, termed** the **Agora**. As we **shall see, it** is not a matter of a complete change from one situation to another but of a growing complexity where key elements from each phase still **fulfil** main functions in the following phases. The arrows **illustrate** these connections.

Figure 1. Research evaluation in time





3. The concept of the Agora

The dominant idea of a linear and almost parallel development of economic growth and scientific knowledge formulated by Vannevar Bush in 1945 has been shown to be increasingly unrealistic and a hindrance to a new understanding of the complexity in the merger of science, scientists and society (Stokes 1997). Bush's central message, that the most efficient way to organise investment in science was to influence the distribution of money but to leave what goes on in science to the scientists, was reinforced by sociological theories explaining and legitimising the division (Merton 1968, 1973), with the CUDOS norms being the most prominent. The idea of a separation or a construction of a Berlin-like wall between science and society is clearly not valid today even if it is still promoted — not least by some scientists.

The concept of the Agora originates from classical Greek history describing the town square where people met to discuss, debate or simply exchange points of view. The concept was introduced by Nowotny, Gibbons and Scott (2001) in order to discuss changes in the relationship between science

and society and to introduce the idea of a new logic and new social relationships between experts, scientific knowledge claims and society. What is new compared to the linear model are the attempts to describe the multitude of interests and participators — in decision-making regarding and influencing the development of science, in debating and discussing the relevance and 'truth' of scientific knowledge, in the demarcation or boundary lines (Gieryn 1983, 1994) between scientific knowledge and other forms of knowledge, and in prioritising resources and other policies towards knowledge.

In fact, the Agora should not be understood as a new physical location or a specific new policy organisation, but as a common name or denominator for people, places, networks and institutions with interest in or influence on the struggles to certify scientific knowledge claims. This means that the Agora contains many of the well-known science institutions: journals, scientific conferences, universities, research councils, ministries, parliaments, media, public opinions, social movements etc. The important change is that one cannot anticipate where the essential discussions and final decisions about knowledge claims will take place or who will be the decisive actors. However, the Agora does not only consist of or is not identical to the institutions — the concept of the Agora stresses the importance of the relationships, the processes and the dialogues taking place between the many different actors — a phenomenon that the authors call 'contextualisation'. It is important to emphasise the difference between the classic concept of the market and the concept of the Agora. The Agora is much broader and encompasses not only the traditional market for products with quantities and prices (which is often implicit when discussing the market), but also the political market for negotiations, the public sphere for recognition and the scientific market for reputation (Foss Hansen, 1989). The concept of the Agora is a challenge to traditional views of the role of scientific experts, trust in scientific knowledge etc.:

'The main challenge in the *agora* comes from a deeper level. It arises from the contextualization itself, which has so greatly enhanced the success of science, opening up new research fronts, expanding research horizons and establishing new and apparently flourishing research fields. But contextualization has a price. Science and scientists have not been used to the context speaking back, so it is not surprising that they see contextualization as a challenge to their cognitive and social authority' (Nowotny et al. 2001, p. 207).

The question of the context 'speaking back' is central to the understanding of the Agora. What we hope to be able to achieve in this paper is to clarify the function and expand the understanding of the Agora by relating it to the development of the field of research evaluation. The field of research evaluation can be seen as an important part of how the contextualisation mentioned above can 'speak back to science' and thus research evaluation can form an operational involvement and participation in the Agora.

However, analysing research evaluation from the viewpoint of the Agora is not neutral and does not leave research evaluation untouched. The concept of the Agora implies changes in all the traditional views upon science and scientific knowledge. Science in the Agora no longer has the same authority to certify knowledge, and research evaluation can therefore not be viewed as based on the opinions of scientific experts. It represents a dramatic shift in the understanding of research evaluation that contradicts its usual and traditional perspectives. It is therefore a step that should not been taken without convincing empirical support.

4. From the 'Royal Geographical Society' to a complex relationship between science and society – a history of research evaluation

In this section will we describe broadly the development of research evaluation as it is usually perceived. We will focus on the growing complexity of the evaluations conducted — especially expressed in the increasing number of criteria involved in evaluations.

Evaluation has been an integral part of the research system since the appearance of the first scientific journals, with peer review as its usual form. It is used in a wide range of selection processes, such as applying for an academic position, the scientific publishing process, applying for research grants, for academic degrees and prizes. An evaluation of research performance is a necessary and very natural element in those processes. Researchers themselves often claim that they are the most frequently evaluated professional group.

Over the last twenty years or so, we have seen a development of new types of evaluation, supplementing the peer review both in terms of scale (amount) and scope (purpose). As noted by the OECD, research evaluation has emerged as a 'rapid growth industry' (OECD 1997, p. 5). This development can be seen — among other factors — as a consequence of the new understanding of science and society (a shift from a mode 1 society to a mode 2 society). When research or knowledge production is developed in a 'context of application' (Gibbons et al. 1994) and in a 'context of implication' (Nowotny et al. 2001), it is expected to contribute to various goals (scientifically and practically). And when knowledge and research are produced by different institutions, it is necessary to develop and use new types of evaluation. The development can also be described as management of research under steady state conditions: 'If there is a single word that epitomises the transition to "steady state" science, it must be "management"' (Ziman 1994, p. 272).

These changes in research evaluation have many sources — from general societal development and from experiences of evaluations in other areas. The important changes in perspectives can be observed in several dimensions. The first 'answer' to the problem of creating societally trustworthy or robust information on scientific quality under changing social conditions came in the early 1980s. As part of the beginning general reforms of the public sector, often known as New Public Management, the 'value for money' or auditing of publicly sponsored research became part of this development. This meant that the traditional evaluation carried out by peer review was modified and supplemented by other forms of evaluation, for example aiming at increasing the effectiveness of scientific organisations. In this way, criteria other than the strictly scientific (e.g. truth, consistency and correctness) were also introduced into the evaluations. The major thread in this development was the creation of a large number of quantitative systems or techniques for counting almost everything, the culmination of a century-long development described by Theodore Porter as 'Trust in numbers' (1995). Michael Power (1997) captured the controlling dimension of this development in the concept of 'the audit society', describing how the visible traditional forms of control in organisations are replaced by new internal systems of control and monitoring. In this process, trust is based on information produced by the control systems and not by some inside information accessible only to the community of scientists. A result of this development is a marked tendency to base trust on the use of quantitative information.

In the light of 'the audit society' one of the purposes of evaluation is, as in all other public areas, to control expenditure or at least to create an impression that expenditure is under control by public managers. Stressing the latter aspect, we can talk about symbolic evaluations. In many settings, evaluations have developed into a kind of ritual behaviour. They do certainly have effects, but not necessarily in a traditional rational manner of measuring official purposes or goals (Dahler-Larsen

2000). Another basic — but not official — purpose of evaluations is (just as the earlier discussions of the utilisation of applied research in general (Weiss & Buculavas 1980, Weiss 1999)) to use them as political ammunition in a decision-making context in order to legitimise decisions (Albæk 1996, Hansson 1998). With this view of evaluations, it is also clear that social and political criteria must be seen as playing an important role in evaluations. Evaluations are still based on scientific expert opinions, but they are also used in and integrated into more political games.

All the changes mentioned above in the understanding of why we evaluate knowledge — from rational explanations and organisational development to ritual and symbolic functions — are not the only changes in the understanding of research evaluation. In the following sections, we will examine some of the other changes and developments. This covers different initiators and users of evaluations as well as different evaluators. Furthermore, it is characteristic for the development of evaluations that they have seen both a change in focus and a large degree of variation of foci regarding purpose, levels of research system and the time dimensions for evaluations.

These changes are not a matter of a sudden and permanent shift from one perspective to another, but rather that the new foci will exist to different degrees in various evaluations. Individual Researchers often perceive modifications in just one of the dimensions as quite a comprehensive change.

To summarise: society no longer trusts a one-dimensional evaluation approach conducted in an isolated system with one-dimensional and intrascientific quality criteria. The increase in the number of research evaluations, the many new forms, and the many new criteria used can be viewed as an attempt to create public trust in research — an attempt that results in a significant increase in complexity.

A major problem when quality control is to take into account the many new and complex relationships between science and society is that scientists have not been used to feedback and influence from other parts of society:

‘Caught on the defensive, they blame contextualization, and the agora in which it emerges, for the rise of anti-science sentiments, for the subversive influence of social scientists and other “relativists”. They fear that irrationality will break through the fragile crust of scientification.’ (Nowotny et al. 2001, p. 207)

Consequently, one of the main challenges of research evaluation today is to re-establish trustworthy relations with research in general as well as to find new specific trustworthy evaluation forms that can reach the various groups that participate in the discussions about scientific knowledge: scientists, politicians, the media, social movements, non-governmental organisations, business people (and lay citizens to include everyone) — forms of evaluation that can integrate even more criteria and still handle the complexity. The question is whether it is possible to do this and still uphold the basic understanding of science and scientific knowledge as based on the expert opinions of scientists. Or if this development demands a fundamental shift in our understanding of science and the relationship between science and society — as epitomised in the concept of the Agora.

5. Different forms of research evaluation

In this and the following sections we will substantiate the broader outline above of the evolution of the field of research evaluation by describing more specific forms of research evaluation and their role in the new and changing role of science in society. With the help of four cases describing what

we consider to be central elements in the development and function of research evaluation today, we aim to produce a more comprehensive empirical description of the research evaluation system and its function in a changing relationship between science and society.

5a: The classic peer review

The peer review system (see the case description in Box 1) has over the past hundred years acquired a reputation as the best method of selecting the highest quality in research. It functions in a somewhat similar way to the legal system. Today, the peer review system is as important as ever in the selection of scientific quality, but it is a system that is not easy to administer in a world where the demands for control, accountability and assessment of research are growing. These make demands on peer reviews on a large scale and tend to shift the peer review system towards an administrative tool for a number of new applications.

The most important prerequisites for a good peer review are motivation, competence and independence among the reviewers (Kostoff 1999, Hemlin 1996). At the same time, however, it is very costly to meet the demands for a good peer review. The cost problems are important too, not only in the form of direct costs but also and perhaps most importantly in the amount of time consumed by peer reviews (Kostoff 1999). In addition to the administrative problems and demands, new knowledge is difficult to evaluate in a world of competing knowledge paradigms and a growing integration of knowledge production in new organisational settings outside the universities (Frederiksen & Husted 2002). Considering these problems, the classic version of the peer review system has been under scrutiny as the major evaluation model in the field of research (Hansson 2002).

To summarise: peer review is the traditional evaluation method in science and is still effective in producing trustworthy and legitimate information about scientific knowledge. It is mainly used in internal scientific evaluations and focuses on individual scientific products. The criteria used in peer reviews are supposed to be cognitive.

Case 1 Peer reviews

In one of the most comprehensive surveys of the many single studies of the peer review process, together with an extended discussion with some critics, Cicchetti (1991) demonstrated that although the documented low reliability of peer reviews in single cases is well known by active researchers, the general level of trust and acceptance of the system is not jeopardised. A large proportion of papers rejected in the first instance are, after re-writing, eventually published by the same or another journal. The actors — editors, reviewers and researchers — know in advance that different reviewers with different paradigmatic backgrounds will act differently in the review process. Cicchetti's own conclusion — greater reliability by better training of reviewers and editors — overlooks the main point highlighted by a number of critics. The process and learning dimensions integrated in the peer review system are an integrated part of the constant changes in the scientific disciplinary and cross-disciplinary fields and paradigms and can only work with a certain level of openness and disagreement because of the basic nature of the creation of knowledge of the unknown. It cannot be 'solved' by technical and institutional means without blocking the very fundamental reason for peer reviews, the evaluation of new knowledge by others. It does not mean that the many documented 'wrongs' in peer reviews, the Matthew effect, 'the old boys network', gender bias and other types of biases should be ignored, but it is unrealistic to attempt to isolate peer reviews from other parts of society. All research evaluation is, to quote Elzinga (1995, p. 19), 'socially constructed through conflict and negotiation over control involving sets of actors in different policy cultures'.

The peer review system is the most important selection mechanism for the publication of scientific knowledge (papers, books) as well as in decision-making concerning research grants and projects. Furthermore, it has over a long period demonstrated its flexibility and ability to select the relevant knowledge in ways that are generally accepted by the scientific community as 'fair'. Large research policy institutions such as the NSF in the United States have experimented with the organisation of the peer review process in order to overcome or at least reduce the influence of some of the inherent problems of this qualitative and subjective evaluation system — problems such as an 'old boys network', transdisciplinary versus disciplinary work, favouritism, lack of reliability and, not least, the growing costs associated with high quality peer reviews.[i]

5b: Modified peer reviews

With the extensive use of the peer review system and the general acceptance of this evaluation system by the scientific community, it is natural to develop this form of evaluation further in order to integrate issues other than the strictly cognitive ones. For example, it is possible to involve extra-scientific issues such as the research organisation and organisational development. The *modified peer review model* is an example of an attempt to modernise the classic peer review model in order to use it as an instrument in decision-making regarding a research organisation's development. It has been developed over a number of years in Denmark and other Nordic countries (see Box 2 for the case). Evaluation of university research resulted from a political demand by the Danish government in the late 1980s, partly due to several OECD studies and publications on the necessity of developing a national state research policy and partly due to the overall implementation of the general concepts of auditing and quality management from New Public Management. The evaluation resulted in the implementation of a model, essentially as the result of compromises between the central government and the universities. The 'compromise' was the 'modified peer review model', an open model for evaluating research in context, where the evaluation criteria, evaluators and specific procedures are open to negotiation. The organisational intervention following this evaluative procedure — with a large degree of success in the form of acceptance by researchers as well as by policy makers in the organisation — is still based on a cognitive framework and the scientific products. However, it also brings in new social and organisational criteria and has a specific focus on group learning.

Case 2 Modified peer reviews

At Copenhagen Business School in the early 1990s, organisational development took the form of an evaluation of research quality and potential in individual departments, based on a selection of publications by the department presented to and evaluated by external peers combined with site visits (Foss Hansen & Borum 1999, 2000). After some years of experience with evaluations conducted after top-down decisions from the Ministry of Education with unfavourable results, universities took over and implemented a new modified form of evaluation aimed at both research quality and organisation, and used the results increasingly as an internal instrument for organisational development. According to Foss Hansen & Borum (1999), the newest universities in Denmark have been remarkably involved in initiating this bottom-up evaluation model, probably in order to legitimise 'their status and position in the academic realm' (ibid. p. 311). However, as often happens with new methods of evaluation after one or two years, this modified model became more or less institutionalised or standardised as the internal organisational standard for research evaluation. An open and less formalised model such as the 'modified peer review model' becomes entangled in procedures and standardisation over time. The modified peer review model is based on a traditional standard [illegible] peer review, and is primarily a standard for organising institutional evaluations. However, a combination of pressures for cost reduction and development of routines as the natural result of a bureaucratically organised activity will after a certain period result in a more formalised and standardised 'modification' of the modified peer review model. For instance, the costly peer reviews tend to be replaced by quantitative techniques, especially bibliometrics (Kostoff 1999, Brunsson 2000).

5c: Integrating economic and political issues

The evaluation activities in the first two cases took place inside the universities as organisations or as a part of the disciplinary system (the invisible college). Evaluatees as well as evaluators belong to the scientific community, and the use of evaluations is a task for the universities too. In this third period, the area where research evaluations are performed is enlarged, taking into account the new aspects or approaches of producing scientific knowledge, labelled mode 2 by Gibbons et al. (1994). Scientific progress and quality is not only viewed as an end in itself, referring only to the closed community of scientists, but also as a means for other societal goals. Society is more interested in what happens in science and at universities in order to secure and promote economic growth and political stability. Furthermore, the forms of evaluation of this period illustrate a development towards a standardisation of evaluation procedures and to a certain extent towards a standardisation of

methods as the result of a more general societal development, where evaluations are used as an instrument to steer and control (Hansson 2002).

Science is part of society and when scientists work to a greater extent with problem-oriented tasks, the focus for evaluation changes from quality per se to relevance — have the problems that gave rise to the research questions been solved etc. The problem and the challenge for society is how can we measure the societal benefits of expenditure on research and the 'value for money'? The approach of using a standardised procedure is often related to evaluative purposes, frequently as an activity labelled 'benchmarking' (see case 3). Originally from business, benchmarking is a tool to improve the function of a system by analysing the best practice of other companies to determine standards of performance. The next step is then to implement procedures to achieve those standards. Principally, therefore, its purpose is learning. However, it is so widely used that in many ways benchmarking appears more to resemble a label than a concept. Often the term is used simply as a synonym for ranking. The results are increasingly connected with decision-making processes. It can, for instance, be at research council level or at national level.

Common to many benchmarking activities is the use of indicators. Indicators are used when it is impossible to measure a construct directly, e.g. productivity, quality, impact on innovation.
A frequent recommendation is that the best method is a combination of peers and indicators. Normally it is stated that indicators can never stand alone, but must be complemented by expert opinions. Van Raan (1998) reverses this and emphasises that particularly in scientific areas with many and equal paradigms it is necessary to secure support from indicators. They could constitute a guarantee for an objective approach. The use of indicators may contribute to a separation from peers in knowledge assessment.

To summarise: the main changes in this form of evaluation are concerned with the purposes of evaluations (focus on political decision input and societal usefulness), methods (use of indicators) and that evaluators as well as users are to a greater extent detached from the university system. The scientific content is largely black-boxed in these forms of evaluation.

Case 3 Economic and political issues in research evaluation
On an *international* level, for several years there have been various reports from the OECD concerning the scientific level in member states. A current example is 'Benchmarking S&T productivity' by an EU expert group (2002). Such a report does not have a direct influence on research policy or behaviour at university level or on funding allocation decisions, but it may have impact on national policies. It may stress specific issues on the agenda, perhaps communicated through the media. At the *national* level, there have in recent years been several examples of a development of standardised procedures for assessment combined with resource allocation.

Probably the best-known example is the UK model — the Research Assessment Exercise (RAE). It has been conducted since 1986, through a relatively standardised (but revised) procedure. The total academic landscape is divided into more than sixty units of assessment, and there has been — over time and between the units of assessment — a variation in the specific methods. In general, it is based on the judgements from peers: some of the units and evaluation rounds place greater weight on bibliometric indicators, but it is essentially a mode 1 activity. The exercise has a number of consequences for universities. The universities know they have a carrot to reach — and they try to do so. There is pressure on individual researchers to focus on publishing, and there is pressure on departments and institutions to focus on publishing at the expense of teaching and links to the surrounding community. Cave et al.(1997, quoted by Guena & Martin 2001, p 8.) found clear evidence of 'publishing inflation', i.e. maximisation of the number of articles produced, either by repetition, a lowering of quality standards, or the 'salami slicing' of research into smallest publishable units. As a consequence, there may be a tendency towards greater similarity between universities. In many ways it seems that the exercise is counterproductive in regard to its purpose. There is little probability of a greater emphasis on societal mode 2 goals. It comes out as the decision makers have faith in the linear model of growth combined with a greater efficiency.

There is an ongoing debate on procedure and method development advocating for use of peer review, questionnaire surveys or indicators. All the above methods should be assessed in regard to their benefits and costs. The first two in particular are intuitively considered quite expensive. Considering these costs, Oppenheim (1997) recently concluded, with respect to the British exercise in evaluation: '…citation counting provides a robust and reliable indicator of the research performance of UK academic departments in a variety of disciplines and…for future Research Assessment Exercises, citation counting should be the primary, but not the only, means of calculating Research Assessment Exercise Scores'. The UK is not the only country to have developed standardised procedures and some use of indicators. In general, indicators as a direct decision support have a relatively limited impact, but they do have some effect on the field.

The evaluation of science and research described so far has definitively moved away from the university regime. It is a result of society's interest in science in a 'steady state' regime. However, it has certainly not reached the wider society. Research evaluation activities have developed as a relationship between authorities and universities, and are neither a result of nor will alone result in a stronger contextualisation of the knowledge production system. It sometimes appears that because of the difficulties of measuring the desirable phenomena, there is a tendency to focus on what the indicators show, i.e. on the figures rather than the underlying processes. A consequence of this can be changes in behaviour, but in the opposite direction than intended. If the new procedures are intended to bring about a closer relationship to the wider society context, it is crucial that indicators as expert knowledge do not have the intention or function of closing down the debate, but rather to open it. If the indicators have been co-produced by the social actors in an open and public way, they can serve as 'socially robust' indicators (Barré 2001). This line of reasoning also leads onto the following section.

6. The new complexity in research evaluation

The three levels of research evaluation described in the cases above will now be viewed together and in combination. Figure 2 below provides a picture of how diverse and complex the field of research evaluation has become today — the result of some twenty years of attempts to influence and direct scientific research and to open up science. The evaluations take place at different organisational levels, ranging from personal to transnational, and the evaluator changes accordingly from a scientific peer to a broad arena of political groups, lay panels etc. The figure creates a very complex picture of research evaluation as a very broad and versatile field of activities, from classic peer reviews to benchmarking, ratings, accreditations and user studies, hearings and conferences. The picture is neither complete nor systematic but aims to give an overview to illustrate how the growing organisational complexity and the professionalism of evaluators produces a large number of new types of research evaluation and assessment systems or procedures to establish 'socially robust knowledge' (Nowotny et al. 2001).

One thing seems to be common to most of the different types of evaluation, namely that it is a one-way street. The impact of the evaluations is a communication or decision about a product, result, organisation etc., but there is seldom real communicative feedback, so it is reasonable to state that a genuine dialogue or an interaction in relation to the substance, e.g. the knowledge production, does not take place. In a certain way, these evaluation systems are products of the linear model and its assumptions of independence between producers and users of scientific knowledge. Only in the last tier, evaluations by laypersons, politicians, NGOs etc., is it possible to find some examples of communicative or negotiating processes based on context and interdependency inside an organisational framework.

Figure 2. THE MANY DIMENSIONS OF RESEARCH EVALUATION[1].

Evaluator ➔ / The evaluated unit ⬇	Scientific Peers	Professional evaluators	Lay persons, NGO, political groups and panel
Individual	**Peer review of articles for publication, appointments, awards**	**Research management, researchers productivity, human resource management**	**Public debates, writing books, Participation in media**
Institute/Department	**Modified peer review Rating/quality assurance Peer advisory board**	**Research Management: - marketing - resource managing - benchmarking**	**User studies (as interest groups in the field), training of young researchers, production of new knowledge (patents, products)**
Institution	**Modified peer review Rating/quality assurance Research advisory board**	**Contract steering- Negotiation of conditions (contracts) Resource allocation Benchmarking TQM**	**User studies (as interest groups in the field), quality, employers of candidates, users of new knowledge (business, politics)**
Cross-institutional Evaluations (institutions, disciplines, areas)	**Informed peer review, Rating/ quality assurance site-visits, Research advisory board, Accreditation; EQUIS, ISO9000**		**Social consequences, Public hearings, consensus conferences, Public or NGO advisory boards**
Territorial/-nation state National Policies	**Rating/Expert-statements, visions, scenarios, proposals. Evaluation and prioritation of areas, cost-benefit studies, forecasting**		**- Political debates, public discussion of visions, public or NGO advisory boards**

[1] **Finn Hansson**, Department of MPP and REMAP. Workshop on Upgrading the Danish University System, CBS September 25.th 2002.

The many dimensions and the complexity in Figure 2 should be understood as the outcome of a vast number of specific attempts to solve the basic problem: how to produce trustworthy or socially robust knowledge regarding research quality for the traditional research community, the political system and the wider public. The trust in research evaluation was once based only on the norms of the scientific community itself through the reviews by peers (Merton 1963). The classic peer review system continues to be important today, particularly for publication. However, it has not been able to provide answers to the questions of accountability and of social and political acceptance and trust in new scientific knowledge claims. These answers are to be found in the many new attempts to introduce users — the public, NGOs, the political system — through a number of new evaluation procedures and forms listed in the right hand column. Alongside the classic system of knowledge control and its many newer variations (the left column), we have the general public discussion on science and assessment methods normally used here (the right column). In between, we have a large number of different and mixed attempts to solve the problem of producing trustworthy knowledge on the value of research towards specific goals.

Evaluation of scientific knowledge has always been a question of how to establish trust in some specific scientific knowledge production. The problem we face today is how to establish trust in scientific knowledge claims in modern societies with many complex relationships between science and society at large, and what kind of procedures and institutions are needed in a situation where the decline of the traditional authority of science occurs hand in hand with a rapidly growing dependency on expertise and science-based knowledge claims (Beck 1999, Nowotny et al. 2001). How can we develop the field of research evaluation to meet the many contradictory demands and situations, and can we point to some new 'rules of the game' defining the roles and interplay between scientists, politicians and laypersons?

7. New combinations – the rise of the Agora?

So far we have discussed the development in research evaluation from an internal scientific view to a much more open and external function and orientation of the whole evaluation process — placing a broader perspective on research evaluation that focuses on science in relation to society. The use of indicators, benchmarking and quality assessment systems in science policy has opened up the former closed research system with its 'own' quality control, the peer review system, to a number of 'outside' influences and controls. However, are we still observing an opening under control in the sense that the access and understanding or reading of the information are restricted to a rather well-defined forum of experts, science policy officers etc.? And can new criteria for handling the interests of the broader lay public be integrated in such a way that the original aim and view of research evaluation can still be maintained? Or is it necessary to change our fundamental perspective on what research and research evaluation is all about? This is the change that the concept of the Agora attempts to encapsulate.

In complex, open, modern society, the many different types of research evaluations are not and cannot be restricted to a closed world of scientists, experts and science policy officers. Public debate has in an increasing number of cases taken up questions in relation to the evaluation or assessment of scientific knowledge and made them part of public discourses in order to influence the outcome, e.g. the decisions on the use or non-use of a certain new technology or other application of new scientific knowledge. However, what research questions are taken up, and by whom, is impossible to predict today. This uncertainty is precisely one of the general characteristics of the new concept of the Agora. Who the specific actors are in a new discourse on a research question and how and why the question is raised are not foreseeable.

Furthermore, the interplay between old and new research evaluation procedures in universities and other research organisations and the assessment of scientific knowledge claims in the public sphere are areas we know very little about. The concept of the Agora should be understood as a demonstration of this lack of knowledge. What we can observe is a growing number of new actors participating in the field. Among the more important new actors are: (a) private organisations (health interest groups, environmental NGOs etc.) who employ scientific expertise in their respective policy field, (b) large companies promoting their own research evaluation of results within their commercial interest sphere,[ii] (c) private research organisations and corporate universities, and (d) a growing interest and professional handling of questions of scientific knowledge by the media. To this is added the open question of who acts in the public sphere in relation to a specific science-based question and how does the public sphere evaluation or assessment of scientific knowledge influence in turn the internal scientific procedures.[iii] These very complex questions are exactly the core agenda of the discussion of the Agora. In the following case, we intend to illustrate the problem of

the many unforeseen participants and actors and the diversity of arguments and methods used to assess scientific knowledge in public debates.[iv]

Case 4: New combinations – the Agora?
Before the summer of 1999, the multinational corporation Monsanto argued for permissions to market their newly patented GMO-based food (corn, soybeans) to the European market. To support their claim they presented to various national and EU agencies a large amount of peer reviewed scientific evidence on the revolutionary effects of GMO in different basic foods and its safety. The EU science advisory system discussed risk and contamination limits in relation to a general licensing of GMO-based food, attempting to assess the long-term environmental effects. However, the various EU agencies had neither the resources nor the access to reproduce or control the knowledge based on laboratory experiments presented by Monsanto, but had to rely on the certification of the results and expert statements produced by the company. At the same time, local politicians and NGOs challenged the safety claims by Monsanto (claims of no risk) with the help of alternative scientific information on the risk issues from university scientists. In June 1999, several EU countries issued a joint declaration stating their intention to 'take steps to ensure that any approval for the sale or release of genetically modified organisms (in the public discussion named GMOs) in Europe will be suspended'. This declaration was the basis of what has become known as the 'de facto moratorium' on genetically modified organisms in Europe. The background to this decision was a long public debate in almost all member states on the possible risks of GMO. The uses of scientific knowledge by both proponents and opponents of gene technology in the context of European agricultural policy heavily influenced the debate.

In the debate on the use of GMO in the EU, proponents deployed scientific knowledge claims based on peer reviews in order to solve the problem of public and political resistance and uncertainty, using arguments such as 'The best scientists in the world have shown these means to be safe',[v] normally combined with (often quite passionate) appeals to the potential benefits of biotechnology, for example '…researchers are working not to boost farm incomes, but to save the lives of their people'.

The opponents, lay Europeans, often formulating their concern through a number of NGOs and at a later stage several European governments, display widespread scepticism about the food they buy. The scepticism has very clearly been demonstrated in several Eurobarometer surveys (INRA 1998, 2000) measuring the attitudes of consumers towards food safety: 'In general, there is a relatively low level of confidence in producers, whether because of the information they provide about the product or, more importantly, because of the fact that they sell products that are profitable, but not safe (INRA 1998). An earlier survey also showed that there was no simple relationship between public scepticism and the public being more or less knowledgeable. For instance, while the public in Denmark is relatively knowledgeable, they are at the same time rather sceptical. A study of the Danish media debate on the GMO discussion in Europe showed that the scepticism or lack of trust in the scientific knowledge-based claims from the proponents was based on a reasonable level of factual knowledge (Hansson & Horst 2000).

The lines of division between opponents and proponents did not run between accepting and rejecting scientific rationality per se, but between providing answers to two rather different questions related to the assessment of the knowledge in question. First, who appeared a trustworthy representative of scientific rationality? And second, how broad was the sphere of application for scientific rationality or the role of the social rationality in the application?[vi]

What does case four tell us about evaluations of knowledge claims and quality in a complex situation with many different stakeholders? The case of GMO policy in the EU presents a picture of the complexity in the Agora as a marketplace for evaluations and decisions on questions of truth claims and trust in scientific knowledge. It contains public discourses and more closed discussion and confrontation between experts and counter-experts over the interpretation of scientific knowledge and its consequences, an increasing use of alternative expert opinion and surveys in public debate by social movements and organisations — all signs of the weakening of the structure of authority within science and a lack of an 'automatic' social and political trust in scientific knowledge. The many participants in the process — the multinational corporations, scientists, scientist-administrators (EU), politicians (national and EU), social movements/grass roots organisations and the media can be understood as separate and interacting actors in the Agora (Nowotny et al. 2001).

The knowledge in question in this case was a mixture of mode 2 knowledge and the limits or borders of scientific rationality for GMO and its applications. At a certain stage in the process it is rea-

sonable to speak of the construction of 'socially robust knowledge' by the opponents, resulting in the EU GMO moratorium. However, the situation was based more on 'our unawareness to know', to paraphrase Beck (1999), than on an accepted and stable body of knowledge. Later the whole scenario changed. Public interest and NGO activities shifted towards other important questions, but Monsanto continued their pressure, presenting a continuous stream of 'new scientific evidence' on the benefits of NGO food to the EU authorities in order to have the moratorium changed,[vii] demonstrating the impact of economic power and resources on the outcome of policy in the Agora.

8. Conclusion – a complex integration of new criteria or the Agora

Discussions of questions relating to science in society have in the last centuries taught us to be aware of a rather strong and latent tendency to isolate and define the problems in this area as something unique and special, as a consequence of the high prestige normally granted to science. Not even the observation on modernity eighty years ago by Max Weber, who discussed the loosening of the spells of science ('Entzauberung', Weber 1920), has had a lasting influence on the discussion of the role of science in society. The trust in science has traditionally been and to a large degree continues to be based on institutions that are attached to the idea of an autonomous and disinterested science (universities and the peer review system). What we have tried to demonstrate in this paper is that the rise of a whole new field of technical research evaluation systems should be understood as part of the modernisation process taking place in all Western societies where the close links between social trust, visibility and accountability in any part of society have also reached science.

If science is to engage in the developing and changing relationships with society and face the financial interests and power games and at the same time retain the public's trust, demands for a radical change of perspective and implementation of new methods or procedures in relation to the evaluation of scientific knowledge is unavoidable.

In our attempt to discuss the new and changed roles of research evaluation in society, we have tried to describe and partly explain how these different forms or types of evaluation gradually change. The four different archetypes of research evaluation, each illustrated with the help of a case, are on the one hand described in a form of ideal type of the historical development, from peer review to benchmarking; on the other hand, however, it is very important for us to emphasise that these different types of research evaluation develop in a complex social and political setting, where new forms develop as an extension of older forms but not as a separate new entity replacing the old. The result is an ever-increasing complexity in the concept of research evaluation in relation to the social and political surroundings or contexts, a situation much like that experienced by other fields of evaluation. Questions about the function and role of the evaluation, nominations of the stakeholders, participation in and learning by evaluations, the role of evaluation in society (empowerment evaluations, democratic evaluations) and not least the professionalism of evaluations are all themes raised in discussions in many other fields of evaluation during the last ten years.[viii]

It is possible to give at least two theoretically different explanations or interpretations of the consequences of the growing complexity in the relationship between research evaluation as an instrument to influence or manage research and society. In both cases the relationships between the two are mediated through negotiations and interpretations. The role of the media is of great importance in this mediation, but as stated at the beginning of the paper, in order to clarify and focus our discussion on the role of research evaluation we have kept the discussion of role of the media to a minimum.

In one interpretation, what we can see is the development and forming of new organisational and political relationships between the different research fields and the social world, the public. The main agenda is how to organise the interpretation of indicators and measurements picturing the development in research. A number of different proposals have recently been put forward — the uses of consensus conferences (Fujigaki & Leydesdorff 2000), ethical committees on research questions (GMO) — all to establish new social organisations in the complex relationships between the social and political world ('the public') and the world of research in order to create common ground for the trustworthy interpretation of indicators and measurements. In this interpretation, the role of the public is integrated or mediated through a number of political and organisational arrangements in order to raise participation through a higher level of public understanding of science (Wynne 1996, OECD 1997, Expert Group EU 2002). However, some kind of hierarchy of trustworthy knowledge claims in favour of scientific institutions is still clearly at work in this scenario.

The other interpretation of the changed role of research evaluation is the Agora — where results, process and participants are a much more open question. It does not mean that the development of new organisational forms (consensus conferences) are unimportant — they can have an important influence on the mediation processes. The difference to the other interpretation is, to quote Nowotny, how 'the contexts speaks back' and how little of the scientific authority remains. The real contextualisation implies a completely new scenario, where results from research evaluation, unlike earlier, are a given input in the negotiation and interpretation but have to fight to obtain social trust. It is not a new agenda for a Habermasian ideal speech communication, because the social and political power structures remain unaffected in the background. What has changed is the agenda setting and interpretation. What the dominant interpretation of the results of the evaluations and indicators will be is not something that can be foreseen. And the same is true for who will be central players or participants. Large corporations or international or government agencies will of course be important and influential in the whole picture, but when analysing individual case situations, such as the GMO case, the point in the concept of the Agora is that nobody can predict or foresee who will set the agenda and what the central problem will be. The concepts of 'post academic science' (Ziman 2001) or 'post normal science' (Rawaetz 1999, Luks 1999) have been proposed to describe this picture of the real demystification of science. The role of the public in the Agora is much more complex and is based on a radical de-monopolisation of knowledge claims by academic and other formerly privileged institutions.[ix]

What interpretation is then the best? The one that argues for viewing the development of research evaluation as a continuation of an ongoing process that constantly increases the number of approaches to research evaluation — giving more complex kinds of evaluations involving an increasing number of criteria? Or the one that emphasises that the appearance of the Agora means a real and radical change in the relationship between science and society and therefore also a change in the status of all the various approaches and elements of research evaluation? We think that the above examples and discussions show that the Agora is at the moment the best possible understanding and explanation of these changes. The traditional perspective can only understand the many mass mediated discussions of the status of scientific knowledge involving the broader lay public if we develop it beyond recognition.

References:

Albæk, Erik (1996). Why all this evaluation? Theoretical notes and empirical observations on the function and growth of evaluation, with Denmark as an illustrative Case. *Canadian Journal of Program Evaluation*, vol. 11, (no. 2), 1-34.

Barré, Rémi (2001). Sense and nonsense of S&T productivity indicators. *Science and Public Policy*, vol. 28, (no. 4), 259-266.

Barré, Rémi (2001). The Agora model of innovation systems. S&T indicators for a democratic knowledge society. *Research Evaluation*, vol. 10, (no. 1), 13-18.

Beck, Ulrich, Scott Lash & Anthony Giddens (1994). *Reflexive modernization. Politics, tradition and aesthetics in the modern social order.* Polity Press, Cambridge.

Beck, Ulrich (1999). *World Risk Society.* Polity Press, Cambridge.

Brunsson, Nils & Bengt Jacobsson (2000). *A World of Standards.* Oxford University Press, Oxford.

Cicchetti, Dominic (1991). The Reliability of Peer Review for Manuscript and Grant Submission. *Behavioral and Brain Sciences*, (no. 14), 119-186.

Clark, Burton R. (1998). *Creating entrepreneurial universities: organizational pathways of transformation.* Pergamon, Oxford.

Dahler-Larsen, Peter (2000). Surviving the Routinization of Evaluation. *Administration & Society*, vol. 32, (no. 1), 70-93.

Elzinga, Aant (1995). Reflections on research evaluation. *Science Studies*, vol. 8, (no. 1), 5-23.

Etzkowitz, Henry & Loet Leydesdorff (2000). The dynamics of innovation: from national systems and 'Mode 2' to a triple helix of university-industry-government relations. *Research Policy*, vol. 29, 109-123.

Etzkowitz, Henry & Loet Leydesdorff (1997). *Universities and the global knowledge economy: a triple helix of university-industry-government relations.* Pinter, London.

Expert Group (EU) (2002). *Benchmarking the promotion of RTD culture and public understanding of Science.* Brussels.

Expert Group (EU) (2002). *Benchmarking the S&T Productivity.* Brussels.

Foss Hansen, Hanne (1989). *Academic autonomy, bureaucracy and market : the development of public control of the universities. The case of Denmark.* COS, Copenhagen.

Foss Hansen, Hanne & Finn Borum (1999). The construction and standardization of evaluation. The case of the Danish university Sector. *Evaluation*, vol. 5, (no. 3), 303-329.

Foss Hansen, Hanne & Finn Borum (2000). The Local Construction and Enactment of Standards for Research Evaluation. The case of the Copenhagen Business School. *Evaluation*, vol. 6, (no. 3), 281-299.

Frederiksen, Lars Frode & Kenneth Husted (2002). Academic publishing activities by Danish companies. A preliminary assessment. *Vest*, vol. 15, (no. 2-3), 9-25.

Fujigaki, Yuko & Loet Leydesdorff (2000). Quality control and validation boundaries in a triple helix of university-industry-government: 'Mode 2' and the future of university research. *Social Science Information*, vol. 39, (no. 4), 635-655.

Geuna, Aldo, Dudi Hidayat & Ben Martin (1999). *Resource allocation and research performance: The assessment of research. Final Report.* SPRU Science and Technology Policy Research, Brighton.

Geuna, Aldo & Ben Martin (2001). *University research evaluation and funding: An international comparison.* SPRU-Science and Technology Policy Research, University of Sussex.

Gibbons, Michael and Bjoern Wittrock (1985). *Science as a commodity. Threats to the open community of scholars conference, 1983, Stockholm.* Harlow.

Gibbons, Michael, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter **Scott &** Martin-Trow (1994). *The new Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies.* Sage Publications, London.

Gieryn, Thomas (1983). Boundary-work and the demarcation of Science from Non-science: Strains and interests in professional ideologies of Scientists. *American Sociological Review*, vol. 48, (December), 781-795.

Gieryn, Thomas (1994). Boundaries of Science. in Sheila Jasanoff, ed. *Handbook of Science and Technology Studies.*(393-443) Sage, Thousand Oaks.

Hansson, Finn (1998). Evaluation Research and Sociology in Denmark. A Tale of Two Cultures. *New Directions for Evaluation*, vol. Spring 1998, (no. 77), 55-71.

Hansson, Finn (2002). How to evaluate and select new scientific knowledge? Taking the social dimension seriously in the evaluation of research quality. *Vest*, vol. 15, (no. 2-3), 27-52.

Hansson, Finn & Horst, Maja (2000). To trump the passions. The use of scientific knowledge in the introduction of new gene technologies in Europe. In Jose Manoel Carvalho de Mello & Henry Etzkowitz, eds. *The endless transition: relations among social, economic and scientific development in a triple helix of university-industry-government relations.* Rio de Janeiro

Hemlin, Sven (1996). Research on Research Evaluation. *Social Epistemology*, vol. 10, (no. 2), 209-250.

House, Ernest R., Carolyn Haug & Nigel Norris (1996). Producing evaluations in a large bureaucracy. *Evaluation*, vol. 2, April, (no. 2), 135-150.

INRA (Europe) European coordination office (2000). *Eurobarometer 52.1. The European and Biotechnology.* Directorate-general for Research, Directorate general B.

INRA (Europe) European Coordination Office (1998). *Eurobarometer 49 - Food Safety.* European Commission DG24.

Kostoff, Ronald N. Handbook of Research Impact Assessment. (1999). Office of Naval Research 1-17 [7th. Edition]. http://www.dtic.mil/dtic/kostoff/

Luks, Fred (1999). Post-normal science and the rhetoric of inquiry: deconstructing normal science? *Futures*, vol. 31 (no. 7), 705-719),

Merton, Robert K. (1968). The Matthew Effect in Science. The reward and communication systems of science are considered. *Science*, vol. 159, (no. 3810), 56-63.

Merton, Robert K. (1973). *Sociology of science. Theoretical and empirical investigations.* Chicago, Ill.

Nowotny, Helga, Michael Gibbons & Peter Scott (2001). *Re-thinking science. Knowledge and the Public in an Age of Uncertainty.* Polity Press, Oxford.

OECD (1997). *The Evaluation of Scientific Research: Selected Experiences.* OECD/GD (97) 194 OECD, Paris.

OECD (1997). *Promoting public understanding of science and technology.* OECD, Paris.

Oppenheim, Charles (1997). The Correlation between Citation Count and the 1992 Research Assessment Exercise Rating for British Research in Genetics, Anatomy and Archaeology. *Journal of Documentation*, vol. 53, (no. 5), 477-487.

Porter, Theodore M. (1995). *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life.* Princeton University Press, Princeton.

Power, Michael (1997). *The Audit Society. Rituals of Verification.* Oxford University Press, Oxford.

Ravetz, J. R. (1999). What is post-normal science. *Futures*, vol. 31, (no. 7), 647-653),

Stokes, Donald E. (1997). *Pasteur's quadrant. Basic science and technological innovation.* The Brookings Institution, Washington D.C.

van Raan, Anthony F. J. (1998). Assessment of social sciences: the use of advanced bibliometric methods as a necessary complement of peer review. *Research Evaluation*, vol. 7, (no. 1), 2-6.

Weber, Max (1992). *Wissenschaft als Beruf 1917/1919. Politik als Beruf 1919.* ed. by Wolfgang J. Mommsen, Wolfgang Schluchter, Birgit Morgenbrod, Mohr, Tübingen.

Weiss, Carol Hirschon & Michael Bucuvalas (1980). *Social Science Research and Decision-making.* New York.

Weiss, Carol Hirschon (1999). The interface between evaluation and public policy. *Evaluation*, vol. 5, (no. 4), 468-486.

Wynne, Brian (1996). Misunderstood misunderstandings: social identities and public uptake of science. in Brian Wynne & Alan Irwin, eds. *Misunderstanding science? The public reconstruction of science and technology.* (19-46) Cambridge University Press, Cambridge.

Yearley, Steven (2000). What does science mean in the 'public understanding of science'. in Dierkes Meinolf & Claudia von Grote, eds. *Between understanding and trust.* (217-236) Harwood Academic, Amsterdam.

Ziman, John (1994). *Prometheus bound. Science in a dynamic steady state.* Cambridge University Press, Cambridge.

Ziman, John (2000). *Real science. What it is, and what it means.* Cambridge University Press, Cambridge.

[i] See e.g. the Report to the National Science Board on the National Science Foundation's Merit Review Process for the Fiscal Year 2001. House, Haug and Norris (1996) have shown how the NSF succeeded in developing an internal high quality evaluation culture.

[ii] On commercialisation of scientific knowledge, see the discussion in Gibbons & Wittrock 1987.

[iii] The problem to be discussed here differs slightly from the discussion that has continued for some years under the label 'the public understanding of science' (Wynne 1996, Yearley 2000).

[iv] The case is based on Hansson & Horst (2000).

[v] In a 1997 speech to the International Grains Council, US Secretary of Agriculture Dan Glickman expressed the 'utmost respect for consumers here in Europe' and their 'healthy scepticism', but he remained firm in his belief that 'sound science must trump passion' in the case of biotechnology, quoted in Hansson & Horst (2000). This can also be taken to mean that in Europe passion has so far trumped sound science, leading to the current EU moratorium on GMOs.

[vi] On the limited ability to predict and control, limited rationality and 'our unawareness to know' in risk society, see Beck 1999.

[vii] At the time of writing (autumn 2002), it appears that the moratorium will be lifted.

[viii] See the themes for the annual conventions of the American Evaluation Association or the European Evaluation Society. The two evaluation journals with close relations to the two organisations *New Development in Evaluation* and *Evaluation* have recently published a number of articles discussing these themes.

[ix] 'To repress or to acknowledge knowledge (on *all* sides), that is, the Hamlet question which is being posed on the dividing line of the second, non-linear modernity. In contrast, a double construction of unawareness characterizes linear modernization. First, *other* forms of knowledge are blocked out and rejected, and, second, we deny our own *inability* to know. This applies not just to experts, but to activist movements as well. The former stands with their backs to the future and operates in the false self-assurance that comes from having denied their unawareness. The latter dogmatize their (un)awareness for purposes of political intervention. It is precisely this admitted uncertainty which opens the context of action for industrial modernity. Both groups would have to look at themselves from outside, so to speak, in order to understand and shape reflexive modernity's horizon of uncertainty in constructive political terms. Both issues of second modernity — the deliberate acknowledge of outside perspectives and rationalities, on the one hand, and the explicit working out and processing of unawareness, on the other — have not really become an issue so far' (Beck 1999, p. 131).