Evaluating Research in Context

A method for comprehensive assessment

Jack Spaapen, Huub Dijstelbloem and Frank Wamelink

Second edition

Evaluating Research in Context

A method for comprehensive assessment

Jack Spaapen, Huub Dijstelbloem and Frank Wamelink

Second edition

The Hague, June 2007

Acknowledgement to the first edition

Contents

Executive summary

"Conventional wisdom is often wrong."
Dr. James Watson, interviewed at the "Winding your way through DNA" symposium at the
University of California San Francisco in 1992.

To develop a new method for assessing the quality and relevance of scientific research for science and society is the topic of this book. This topic is getting ample attention these days, not only in the Netherlands, but internationally as well. Though the question is relevant for most scientific research, it is perhaps more obvious in some cases than in others. Our approach is based on two fields we studied, agricultural sciences and pharmaceutical sciences.

We start from the assumption that research programs develop over time in mutual transactions with the context that is relevant for their work. The success of a research program therefore depends partially on the ways in which researchers manage to connect to themes in that environment, and on the ways in which this environment absorbs ('uses') and further develops the results of the research. Researchers implicate questions and problems raised by their context in their programs. Some research programs develop primarily in connection with the international scientific community (often a disciplinary community), others are more oriented towards European networks in which general policy questions are at stake, or collaborate with professional entities in a context of application. As a consequence, traditional qualifications for research groups such as 'applied', 'fundamental' and 'pure' do not fit the wide variety that can currently be detected.

To do justice to the diversity that has developed in terms of the dynamics and organization of knowledge production we distinguish a number of social domains in which researchers operate. These domains and the variety of interaction between researchers and these domains form the basis for our method. Domains can vary for different fields, but the main ones consist of the international scientific community, industry, politics, the public sector and the general public (end users). In each domain, we develop a variety of criteria and indicators, and bring them together in a single graphic representation. That provides a comprehensive visual depiction of a research group's profile. We call this a Research Embedment and Performance Profile (REPP). The REPP is a quantative reconstruction of the group's activities and performance in a relevant environment. We combine this with three more qualitative elements:
a.   analysis of the mission and/or profile of the research program or group;
b.   stakeholder analysis and
c.   feed back and strategic discussion.

We think that the four steps in our methodology are necessary building blocks for every evaluation procedure that recognizes the contextualized position of research. The exact execution and operationalization of our method may vary (though within a certain bandwidth), depending on the specific local circumstances, demands and constraints. But the following elements are crucial:

1. **Mission:** a phase in which the mission of a group/program and/or its self image is established;
2. **REPP:** a phase in which a more or less objective (quantitative) picture of the group's production and interaction with the environment is established;
3. **Stakeholders:** a phase in which the environment is consulted about the impact of the group's work;
4. **Feedback:** a phase in which the results of phase 2 and 3 are confronted with phase 1, and which is meant to organize a debate on the strategy of the group.

To be short, we analyze the network around the research programs, represented in a distribution of various social domains. Analytically, it implies a combination of two different approaches. Where a domain indicates a certain well defined field or practice with its own norms, work and responsibilities, a network implies that the boundaries between domains are 'blurring'. By making use of those two perspectives we (re)construct a comprehensive image of the field of research. As we explain in the new introduction of this second edition, we will further explore our and other methods to evaluate the societal value of research in a national project ERiC (see appendix 5). In this project the major Dutch science policy organisations work together.

Preface

In this book, sci-Quest[1] presents a method for the evaluation of research in the context of policy and societal questions. This method is developed in response to a specific request of the Consultative Committee of Sector Councils for research and development (COS) that has been formulated in relation to the so-called MKO-project.[2]

The COS has a long-standing interest in research evaluation that addresses a wider context than that of the so-called scientific value of research. That wider context is the dynamic societal environment of research in which a variety of stakeholders operate, ranging from government to industry to (end) users. The sector councils operate in the space between research institutions and these stakeholders, exploring medium-term and long-term scientific and social trends and bridging gaps between the different interests, goals and demands. For more than a decade they have conducted and sponsored studies and conferences in this area.[3]

This publication is an outgrowth of two methodological studies commissioned by the COS and supported in various ways by several other organizations (VSNU, Wageningen University, NRLO, Utrecht University, Groningen University).[4] The studies were conducted between 1998 and 2002, and were to provide external evaluation committees with supplemental information with regard to the assessment of agricultural and pharmaceutical research. The assessments took place in the context of the regular national research evaluations, first under the VSNU protocol[5], then under the recently introduced Standard Evaluation Protocol (SEP 2003[6]). They covered some thirty research programs in total, twelve in the agricultural sciences in Wageningen University, the rest in two pharmaceutical faculties in the universities of Utrecht and Groningen. sci_Quest conducted these studies as part of a larger COS effort to develop a method for the assessment of what is referred to as the 'societal quality' of research. The main goal was to provide insight and systematic information about the way research was performed in a social context in which a variety of stakeholders operate. These stakeholders (government, industry, NGOs, professional and client organizations, the public at large) all in their own way affect the research process and the research agenda, ranging from particular administrative demands to various forms collaboration. Additionally, they sometimes operate as 'co-producers' of science.

The context of stakeholders related to any given research group is unique, as is its impact on the research process. It can take on different forms and have different weight. Its impact can be indirect, such as through public debate about the implications of scientific research (the polypill for

---

1   Sci_Quest is an independent network of researchers and science policy makers in the Netherlands.
2   MKO stands for 'Maatschappelijke Kwaliteit van Onderzoek', in English: societal quality of research.
3   We discussed a first set of COS-sponsored experiments in Spaapen and Sylvain 1993.
4   Spaapen and Wamelink 1999 (available on http://www.minocw.nl/cos/publicaties/index.html) and sci_Quest 2003.
5   Protocol 1994, VSNU, Utrecht, code: PU 13/03/10.
6   Standard Evaluation Protocol 2003-2009 for public research in the Netherlands, KNAW, NWO, VSNU.

everybody over 55 to give an example) or more direct, as when academic researchers collaborate with industrial researchers. The analysis of such relations between research and its stakeholders is at the heart of our attempts to develop an evaluation method. On the one hand, we aim at representing the interactions between a research group and its stakeholder environment through indicators in a graphic way (the REPP), on the other hand, we flesh out that relation in a more qualitative manner through an analysis of the stakeholder environment.

A more practical goal of our project is to help evaluation committees gain insight into the research programs' broad context that is relevant for their missions. The results of our studies were given to the evaluation committees and helped their members assess the research programs in a more comprehensive way than was previously the case. The value of these reports for the committees was evaluated afterwards. Overall they viewed our studies as well-organized and convincing representations of each group's activities and performance. We have used their and other more critical responses in this report to improve the method and some of our recommendations.

While based on an analysis of our two studies, this publication has a more general audience in mind. The emphasis here no longer rests on an analysis of the different research programs, but on a critical explanation of the overall approach, its development over time and the problems we encountered. In so far as specific research programs are presented here, they serve as examples to make a more general point.

We hope that this publication will help clarify how to approach the evaluation of scientific research in the context of societal demands and that it will be of use in more general discussions at the level of science and innovation policy-making. It is not meant as a recipe for success. We can discuss the method's feasibility, we can examine its data collection and its respective methodologies. But that is not the end of it. It is at least as important for all the concerned parties to acknowledge the premise that stakeholders in the relevant contexts of research programs have an influence that has to be accounted for in the evaluation procedure. We will continue our efforts to further improve the methods to evaluate the societal quality of research, now in the ERiC project (see appendix 5).

Throughout the process of bringing this project to completion, the authors have profited from the support and comments of many people. It is impossible to thank everybody individually, but most prominent were members of the Context group, the COS and its various associated councils, members of the evaluation committees and faculty members. The comments they offered were of great value, both in terms of the study's form and content. Special thanks goes to Lissa Roberts, who edited our English in both eidtions, and improved the book's clarity in other ways too. This notwithstanding, the authors remain fully responsible for what comes next.

1 Introduction to the second edition

In this book we present a comprehensive approach to research evaluation. That is, we aim at including in the evaluation process all relevant activities that researchers engage in when doing research. Thus, we focus on the question of how to assess scientific research, not only in terms of its so-called 'scientific' quality, but also with regard to its value for society. To do that, we have to connect political and intellectual developments in science and society with new ideas and methods about evaluation. Presently, this realm is dominated by debates about the way scientific research functions (or is supposed to function) in the knowledge economy or broader, the knowledge society. For one thing, these debates reveal that there are different implications for different research areas. Arguably, research in the humanities has more difficulties to prove its value in the context of the knowledge economy than, for example, research in nanotechnology. This is partly caused by the fact that, despite the broad societal context of the debates, the focus in most evaluation procedures is still on scientific publications in high impact journals, which are more customary in some areas than in other. In this book, we propose a more complete and balanced method to evaluate research, a method that covers not only the different aspects of research production vis-à-vis different societal domains but also the various interactions researchers engage in with social actors from policy, industry and society at large. With this method we are able to more adequately represent the variety of contributions that research can make to the development of society.

This issue clearly also has supranational implications. Though these surpass the scope of this book, we intend to further the international debate about this topic, if only for the fact that both science and society increasingly develop in an international context. But also because, as we will show later in this introduction, the interest in this topic is mounting rapidly in many countries over the last years. In the Netherlands all the major science policy organisations teamed up in 2006 to start a project that is referred to as 'Evaluating Research in Context' (ERiC). This project addresses both the debate about and the methodological development of evaluations that review research in a wider perspective, and it does so with an expanding European and international participation. We will briefly explain this initiative at the end of this section.

The decision to produce a second edition of this book and to develop the ERiC project was reinforced by the positive responses our initial efforts received. Our aims to operationalize a more comprehensive concept of research quality and to implement that concept methodologically in the reigning Dutch evaluation system drew quite some attention nationally. The fact that we experimented in two different fields and in a number of different universities helped. (One was conducted in 1998 and contained an evaluation of Agricultural Sciences at the Wageningen University. A second, conducted in 2002, contained an evaluation of the Pharmaceutical Sciences at the Universities of Utrecht and Groningen). Also, we were able to improve and simplify the

method in the course of time. With these experiments and consequential advancements we aimed at designing a method that would do justice to the variety of research in different disciplines or research fields, as well as to the variety of contextual interactions between researchers and stakeholders in their environment. The comprehensive method that we propose takes into account the fact that most current research is produced in a complex socio-economic context in which demands are made by a variety of social actors. Moreover, research that addresses complex questions (for example aids, global warming, migration, cultural identity) is often multi-, inter- and/or transdisciplinary and is conducted in a context in which experts with different backgrounds, knowledge and expertise operate and different demands and interests have to be negotiated. Throughout this book we refer to this kind of research as MIT-research (multi-, inter- and transdisciplinary research).[7]

This complexity requires a different approach to evaluation than the traditional peer review that mainly emphasizes scientific excellence and relies on publications in high impact journals for its primary indicators. Since quality in our approach is defined as a multidimensional concept which includes the expertise of stakeholders in different social domains, we elaborate on this concept by looking at these different dimensions, distinguishing in each the modes of production and interaction of researchers and a variety of stakeholders. We call this approach to evaluation evaluating research in context.

The two studies mentioned above provide our main reference material. The methodological changes we implemented in the second study (concerning pharmaceutical research) will be used to illustrate more general points (see especially part three of this book, 'The Method'), but these alterations do not give us a 'blueprint' for an appropriate method for research evaluation. It is important to stress that we see our method as a step in a process of further development, both in terms of methodological improvement and in terms of stimulating awareness in the research and the science policy community. This is precisely the reason that we engage in the ERiC project.

Since the two studies we use as a base were published separately (Spaapen and Wamelink 1999), we will not go into their specific results in this book. We will, however, elaborate on the theoretical background of our approach regarding both 'knowledge production' and the broader concept of research, as well as evaluation as an activity which includes users and stakeholders. Our approach is essentially mission-oriented; that is, it starts from the mission of any given research unit and then focuses empirically on activities undertaken to fulfill this mission. Finally, it brings these elements together in a feedback process in which involved actors reflect on the extent to which the mission is realized. Thus, we come to a four step approach: (i) reflection on the mission of the research unit (or self-evaluation), (ii) empirical reconstruction of the research unit and its relevant context, (iii) stakeholder analysis and (iv) feedback.

---

7   In part 1 of this book we elaborate on this concept of MIT research.

Evaluating Research in Context (ERiC)

Comprehensive research evaluation, including questions about 'societal quality' and 'valorisation', requires a broad discussion and an approach with a wide participation of disciplines and other stakeholders. Such evaluation also touches upon governance arrangements between research institutions and government. 'Evaluating Research in Context' (ERiC) is a Dutch initiative that aims at addressing both methodological questions with respect to new forms of evaluation and questions that regard the implementation of these methods into national or even international evaluation systems. In this new introduction we relate our book to this initiative, present its objectives, briefly discuss one of the more recent examples of an experimental evaluation (in the sector of professional education) and reflect on various governance arrangements.

Since the publication of the first edition of this book in 2005 and thanks to the positive response the book received, we have been able, together with the COS, to bring together the major organisations in Dutch science policy who now feel that the time has come to take a major step forward regarding this topic, both in the sense of methodological progress and in generating more (inter)national attention. Clearly, the book itself is not responsible for this development. It did however connect to a number of developments that helped to enable an easy landing. The Royal Academy, for example, had been discussing the topic of evaluation of societal quality on a number of occasions, and had published a number of reports on the subject through its medical council and its humanities and social sciences councils.[8] NWO, the Dutch research council, has generally put the societal impact of research high on its agenda in recent years, both for the organisation as a whole and through its various departments that deal with different scientific fields. Moreover, NWO has been very active in the movement to develop an 'open' strategy for innovation under the aegis of the Netherlands' 'Innovation Platform'. Essential for 'open innovation', the concept that was coined by Chesbrough in 2003 (though it had been around in the literature since the mid-1990s), is the idea that innovation is very much a joint effort of experts and knowledge producers in different organisations. A key factor in success is therefore to find a method to bring these different expertises together in a fruitful way.[9]

The growing interest within these science policy organisations for the societal orientation of research facilitated the joining of forces in an executive group called the Context Group (CG) that, together with the aid of an international expert group, aims at furthering the execution of the ERiC-project.[10] The CG put together an action program that contains a number of initiatives, including seminars for the Dutch universities and *hogescholen* (or universities of applied sciences)[11]

---

8   Council for the Medical Sciences (2002) and Social Sciences Council and Council for the Humanities (2005).

9   Interestingly, Chesbrough identifies as one of the prime factors that caused the old adage of closed innovation (that is inside an organisation) vanished was the growing social mobility of engineers and technologies.

10   The organisations are the Consultative Committee of the Sector Councils (COS), the Royal Netherlands Academy of Arts and Sciences (KNAW), the Association of Dutch Universities (VSNU), the Netherlands Organisation for Scientific Research (NWO) and Quality Assurance Netherlands Universities (QANU). Also, the Netherlands Association of Universities of Applied Sciences (HBO-raad) and the Rathenau Institute, in particular the Science Systems assessment unit, are involved in the initiative. Further information can be found on http://www.eric-project.nl.

11   The Netherlands has a binary system of higher education: research oriented education, offered by universities, and higher professional education and applied sciences offered by *hogescholen*. The official English appellation is universities of applied sciences.

professional schools, an international workshop on 9 November 2007 and a study into the possibilities of a national expertise center. The CG also developed a manual with guidelines for organisations that plan to evaluate their research in the context of societal and policy questions. On its website, a review of current world-wide developments is published, which will be updated regularly.[12]

In the following paragraphs, we discuss the changing ideas about the relation between science and society and the consequences of their changing relation for evaluation. Also, we go into the implications of these new forms of evaluation for policy and, finally, we report on a recent pilot evaluation in the sector of professional education.

1.1 From 'science' to 'research'

Questions regarding the socio-economic and cultural relevance of scientific research have been on the science policy agenda for decades.[13] A common historical reference here is the 1945 Vannevar Bush report *Science, The Endless Frontier*, in which the interaction between science, industry and the polity that grew during WWII was used to underpin a plea for the establishment of the NSF with the understanding that such an organization would tighten the relations between the three major societal domains. The report can be seen as a prelude to the emerging 'knowledge industry' in the post-war era, first and foremost in terms of a military-industrial complex (in the USA), but also in other societal areas such as agriculture and health care.

The coming of age of a knowledge industry in Europe was noticed as well (Drucker 1969), leading to a view of daily life in European society as increasingly affected by scientific and technological knowledge. One of the effects of the growing knowledge industry has been a broadening distribution of knowledge to and within a number of social spheres, resulting in the sphere of academic (university) science losing its exclusive right to the pursuit of knowledge. It also led to more complex national research systems in which other institutional forms, such as large programs and national centers, bridged the gap between (traditional) knowledge production and society (Rip 2003). Geuna (1999) gives four interrelated factors for the intensification of interactions between university and industry, which arguably also apply to society at large:
1. the development of sciences such as molecular biology, material science and computer science, characterized by high levels of applicability and shorter time between the phase of exploratory research and the possibility of industrial development, spur increased interaction between industry and university;

---

12   See appendix 5 for the Action Program and the list of members of the CG and the international expert group.
13   And even for centuries if one reviews the ways in which colonial governments used science and technology as an instrument to reach certain policy and socio-economic goals. See for instance Shinn, Spaapen and Krishna 1995.

2. an increasing budgetary stringency forces universities to seek external sources of income, thereby encouraging them to carry out research work financed by industry;
3. the growing scientific and technological content of industrial production and certain forms of services such as health care make university knowledge more valuable to industry;
4. policies aimed at raising the economic returns of public financed research stimulate the interaction between university and industry with the goal of increasing the transfer of knowledge from the university.

According to Latour, these developments have led to a shift from 'science' to 'research' (Latour 1998). One critical difference between science and research is its relation to society. Science has traditionally been considered as an academic enterprise clearly separated from society, based on the argument that the former's task was to deal with facts, while the latter was concerned with values. Latour illuminates this with an example drawn from history. In one palace, Galileo dealt with the fate of falling bodies while, in another palace, princes, cardinals and philosophers dealt with the fate of human souls. Society creates culture and social relations, in other words, while science studies the nature of things. Whether it was ever truly reflective of the complex relations between science and society, this image certainly fails to grasp the relations we currently see between the work that is done in laboratories, new insights on the nature of human beings, public debates and societal concerns on the risks of modern technologies.

Latour illustrates this new relation between research and society with a beautiful example:

In early December 1997, a group of patients assembled in the AFM (the French association for the treatment of muscular distrophy) raised, through a television campaign (the Telethon) $80m for their charity. Since the disease that triggers the handicap has a genetic origin, for fifteen years now, AFM has invested massively in molecular biology. To the great surprise of the French scientific institutions, for a while this charity funded more basic research on the human genome than the French State! And they developed original ways to map chromosomes that went so far and so fast that they published in *Nature* some of the first maps of the genome – beating, they boast, even the Americans. Then, once this was done, they disbanded the laboratories they had built for mapping chromosomes, and turned all their efforts to exploring genetic therapy, even though it might be a very long and risky shot. The very building of the AFM (at Ivry, south of Paris) illustrates the limit of a metaphor that would disconnect a Science from a society: on the first floor, patients in wheelchairs; on the next floor, laboratories; on the third, administration; everywhere, posters for the next Telethon and donators visiting the premises. Where is the science? Where is the society? They are now entangled to the point where they cannot be taken apart any longer.[14]

---

14   Science Vol 280, Issue 5361, 208-209, 10 April 1998.

We can see this as an example of the fact that relations between academia and the surrounding society have become less distant in the second half of the twentieth century, a process that was accelerated in different times by different political and economic circumstances. In the 1970s and beyond, the relationship grew closer because most governments felt the need for more accountability in the light of decreasing public budgets (oil crises, etc.) and an expanding research and higher education sector (baby boomers). Globalization and more intense international competition in the 1980s and 90s forced national governments to step up their control in the academic sector. This led to a chain of events and trends such as major funding agencies choosing to stress the importance of socio-economic goals in their funding schemes. At the same time, this created room for alternative sponsoring, from industry and societal organizations, which then fostered a debate about the independence of (academic) science.

In the 21st century, the interaction between science and society is ever increasing in the light of the speeding up of the global economy. In Europe, the EU countries have made a number of ambitious arrangements in Lisbon (2000, 'Agenda for the knowledge economy') and Barcelona (2002, the demand that countries spend 3% of their GNP for R&D) in the hope of becoming the most competitive economy in the world.[15] On a wider scale, we have seen the emerging theme of *global sustainability,* for which the interaction of science and society is indispensable. This theme comes to the fore most notably through the *Millennium development goals* of the United Nations, its consequential 'global' research agenda and, more recently, the alarming report of the UN IPCC report global warming.[16]

Policy initiatives in various countries

Following these international developments, many individual countries are paying growing attention to the relations between science and society, particularly with regard to the impact of research on social and technological innovation. In the Netherlands, the cabinet established in 2003 the so-called Innovation Platform under the guidance of the prime minister. Through this instrument, a direct answer to the Lisbon and Barcelona agreements, the government aims at strengthening the innovative force of the Netherlands and ultimately at becoming one of the forerunners in Europe by 2010. Members of the platform and of its various task forces were recruited from the government, the business community, social organizations and knowledge institutions.

One of the policy issues that led to the establishment of the Innovation Platform was the so-called

---

15   In a recent advice to the EU Commission, the Aho Committee notices the gap between rhetoric on the knowledge society and the willingness to take concrete steps to make it [the knowledge society] happen. It is suggested that the EU establishes a Research and Innovation Platform which includes R&D performers, regulators, users and sectoral stakeholders. ('Creating an Innovative Europe', 2006, p. 8).

16   See The Secretary-General's report 2000 for the Millennium development Goals; and for the most recent IPCC report: Climate Change 2007: Impacts, Adaptation and Vulnerability, WG II contribution to the 4th assessment report of the Intergovernmental Panel on Climate Change of the United Nations, April 2007. See also for example Jane Lubchenco's pledge for a new 'contract' between science and society in 'Entering the century of the environment. A new social contract for science', Presidential Address at the Annual Meeting of the American Association of the Advancement of Science, 15 February 1997. She states that "[a]s the magnitude of human impacts on the ecological systems of the planet becomes apparent, there is increased realization of the intimate connections between these systems and human health, the economy, social justice, and national security. The concept of what constitutes 'the environment' is changing rapidly. Urgent and unprecedented environmental and social changes challenge scientists to define a new social contract."

knowledge paradox. Dutch science, overall, has long been able to maintain a very good reputation and position in the world, but somehow it is not very successful in transferring knowledge into practical applications. In a report of the Central Planning Office (CPB 2003) it is concluded that this écart is caused, among other things, by a big difference between the specializations and knowledge agendas of scientific institutions on one hand and both industry and society on the other. For some time now, the attunement of these different agendas has been one of the main issues in policy discussion, witness recent debates in the Dutch Parliament about the impact of research and the ways to make it visible.[17] These disputes not only cover the question of the relation between science and industry, but have come to entail socio-economic progress in general.

In order to bridge the apparent gap we are seeing a growth in studies that address this issue in the light of evaluation. That is, they examine how different evaluation methods might help assess the impact of research on society in a way that enhances the efficacy of science to achieve the ambitions of governments vis-à-vis the development of the knowledge economy. We could give a plethora of examples, but will mention just a few that together show the variety of issues that can be and are being addressed. More information about these examples will be made available on the ERiC-website.

In the UK most research councils are putting a lot of effort in the development of methods to evaluate the impact of research on society and policy. Over the last couple of years workshops have been organised and reports have been written by for example the Arts and Humanities Research Council and the Economic and Social Research Council. In March 2007 the ESRC had a workshop discussing impact methods that had been tried in practice. One of the methods discussed was the so-called Payback method, an approach that is also tried out in the Netherlands by ZON/MW in the health research sector. While the origins of the Payback method lay in financial spheres, the approach that is now being developed for research considers non-financial aspects such as benefits to the health sector, capacity building and informing policy. In Germany, a number of attempts have been made to develop alternative evaluation methods, particularly regarding transdisciplinary research.[18] One prime example is the evaluation of a large project that started in 2000 in the area of social ecology. The Bundesministerium für Bildung und Forschung (BMBF) supported a large transdisciplinary research program aimed at studying social-ecological transformation processes and their implications for society and policy. It set up an extended review process that was to guide and assess the progress of this project.[19] Also in Germany, a group of researchers at the Institut für sozial-ökologische Forschung (ISOE) in Frankfurt am Main is working on the development of criteria to evaluate transdisciplinary research projects. A group of researchers in Spain is conducting a

---

17   Report Dutch Parliament (Verslag TK 29 338).
18   In the project ERiC we produced an extended review of these kind of studies, published on the website in May 2007.
19   Sozial-ökologische Transformationen im Raum - Synthese von raum- und regionalbezogenem Wissen (STRARE)" in collaboration with the Austrian Forschungsprogramm „Kulturlandschaftsforschung" (KLF) 4 February 2002; see also http://www.gsf.de/ptukf, Forderschwerpunkt sozial-Oekologische Forschung.

number of projects geared toward the construction and validation of a methodology for integral assessment of research in the social sciences and humanities.[20] Also, the number of conferences where the evaluation of the societal impact of research is a central issue is gradually growing. In particular, the Austrian science foundation is very active here, and in America the National Institutes of Health (NIH).[21] The debate is also taking place in other parts of the world. In 2005 the Australian government issued a study into a "greater understanding of how benefits from publicly funded research can be realised and [how] ….practical metrics [can be developed] by which returns can be assessed (Allen Consulting Group 2005)." In the study, four broad research areas that society values are distinguished: (1) material (goods and services); (2) human (health, quality of life); (3) environmental (quality of physical environment) and (4) social (social attachments, political rights, security, engaging society). But while they conclude that a range of metrics exist to measure the extent of these values, they simultaneously conclude that governments in many countries (New Zealand, USA, UK, Canada) do not systematically measure the impact of publicly funded research in these terms. In a study by the Canadian Federation for the Humanities and the Social Sciences, notably titled 'Measuring the 'unmeasurable', a different side of this debate comes to the fore. Here, the central worry is not so much the abyss between science and industry, but almost on the contrary, the fact that the position of the 'liberal arts' has become eccentric in a society that focuses so much on the practical application of knowledge (Grosjean *et al.* 2000). These examples teach us that the debate, on one hand, is about the differences in relationships between science and society with respect to various scientific areas (natural sciences, health, humanities, social sciences) and, on the other, the question is how to measure impact. Can we measure the 'unmeasurable'? Do we use instruments appropriate for the job? Have such instruments been developed? These are the generic questions behind the subject of this book. Although we focus here on methodological questions considering evaluation, we inevitably are confronted with the broader issues that regard the context in which these evaluations take place.

1.2 The rise of alternative evaluation methods

The developments described above have major consequences for the topic we are concerned with in this book. Traditional ways of evaluating research and allocating prestige and money are consequently not adequate for dealing with the array of questions that emerge around this new way of knowledge production.[22] In addition to the peer review system, new evaluative procedures have

---

20   Universidad del País Vasco, UPV/EHU, catedra Miguel Sanchez Mazas, Contact: Julieta Barrenechea.

21   The Austrian Wissenschaftsfonds (FWF), organised a conference in April
2006 on New Frontiers in Evaluation and another one in May 2007 (Rethinking the impact of basic research on society and the economy); further information on the FWF website; the NIH in the USA organised together with the National Cancer Institutes a conference on 'team science'in November 2006. See for proceedings at http://dccps.nci.nih.gov/brp/index.html. See also Stokols et al 2003 and Stokols 2006.

22   Both the Council for Medical Sciences (2001) and the Social Sciences Council and the Humanities Council (2005) of the Royal Netherlands Academy of Arts and Sciences are working on new approaches to evaluating research in the context of broader societal questions. Also, ZonMw and the National Health Research Council (RGO) are working together on a system for the evaluation of societal quality of health research (RGO 2007).

developed in which various bureaucracies gained importance over scientific constituencies. Much to the regret of many scientists, major decisions about funding are currently made in clubs of mixed composition and interest, in particular at the national and European level.

Policy developments in the Netherlands

In the Netherlands, this shift of power became apparent in the mid-eighties of the previous century when the system of Conditional Financing (CF) was introduced in the Dutch universities. It was the first major attempt by the Dutch government to intervene in academic research policy by trying to render universities more efficient and relevant. By doing so, the government merely followed a trend that went through most of Europe in the 1970s and 80s. Although most researchers loathed the system,[23] one of its important innovations was to make room for criteria beside so-called 'scientific quality'.[24]

Although not defined in a very clear way, assessment committees were left with enough freedom to use criteria of social and economic relevance. Also, the composition of the committees allowed for inclusion of non-scientists.[25] This was particularly important for fields and disciplines that were connected to some kind of social practice, for example in the areas of medicine and technology. These fields had complained that 'traditional' assessment committees were not able to judge their research properly because of their bias towards international publications as a main criterion. In the new system, the value of research for society was made an explicit criterion. Unfortunately this happened without clear instructions as to how to assess this. As a result, most evaluation committees ignored it. The overall success of the system was meager in any event, due to the ability of the universities – centuries old institutions as they are – to adapt to external threats without having to change too much internally (see Blume and Spaapen 1988 and Spaapen 1995: 76).

Later, when the 1994 VSNU-protocol was introduced in the Dutch universities, the criterion of social relevance was maintained, but again without proper reference or guidelines as to how to apply it. The effect was perhaps worse than in the conditional finance system, because by that time both the universities and the disciplines had learned from the previous experience and mastered a firm grip on these evaluations. In a number of cases, nonetheless, some sort of societal quality evaluation took place, but separated from the so-called scientific quality evaluation (in pharmacy, for example[26]).

---

23   Researchers not only complained about the 'bureaucratic overload' of the system, they were especially worried about its potential threat that bureaucratic criteria would outweigh scientific value and lead to re-allocation between fields for 'the wrong reasons'. See Blume and Spaapen 1988.

24   The protesting researchers were supported by the boards of the universities who did not like the government barging into their affairs, and also by the Social Sciences Council (Sociaal Wetenschappelijke Raad (SWR)) whose members were worried about the growing interference of the state in science, see also SWR 1983.

25   This happened mostly in committees that operated in the technical, medical and agricultural sciences. In other fields, committees did not know what to do with the criterion of social relevance, see Spaapen 1995: 52.

26   Evaluation of Pharmaceutical Research at University of Groningen and university of Utrecht, VSNU, July 1997.

In the Standard Evaluation Protocol (SEP) that replaced the 1994 protocol in 2003,[27] societal quality is mentioned not only explicitly as one of the main criteria ("technical and socio-economic impact, in relation to important developments or questions in society at large"), it also appears in the checklist for evaluators so that they cannot easily ignore it. Furthermore, the SEP clearly states that the evaluation committees should be sufficiently broad with regard to the background of their members. However, no further guidelines are given on how to do that.

The results of evaluations under the new protocol are reviewed on a yearly basis by a committee of researchers. The first report became public early in 2007. Preliminary results show that interest in the topic of societal evaluation is picking up in the universities, but also that there is a lack of understanding of how to handle the topic. There is also a growing interest in the wider academic world, for example by scientific councils of the Royal Academy (KNAW) and in the Netherland Organisation for Scientific Research (NWO), particularly in the area of Health Sciences.[28]

One effect that is particularly important for the topic of this book, is a growing interest in methods of evaluation that focus on (or include) the societal relevance of research. In the Netherlands, such interest is manifest, for example, in political discussions in Parliament following various reports and publications regarding the socio-economic value of academic research.[29] It also appears inside the academic institutions in the form of growing demand for alternative methods for evaluating scientific research.[30]

Research communities dissatisfied

To a certain extent, interest in new methods within the scientific community stems from a feeling of dissatisfaction. Many researchers feel the need to use evaluation methods that do more justice to the variegated character of their work than do traditional methods that basically look at the research production in terms of articles in (high impact) scientific journals. Without a doubt, publication of results in *peer reviewed* journals is important for (nearly) all research areas. But for many researchers, it is also important to collaborate and communicate directly with other groups relevant to their work. In general these groups are referred to as stakeholders and they

---

27   The new protocol came after a tripartite committee wrote a report about the problems that were encountered with evaluation according to the old protocol, such as problems with the assessment of MIT research, too much bureaucracy and not enough research policy consequences of the evaluations. See *Kwaliteit Verplicht. Naar een nieuw stelsel van kwaliteitszorg voor het wetenschappelijk onderzoek*, KNAW, VSNU, NWO, 2001.

28   Both the Council for Medical Sciences (2001) and the Social Sciences Council and the Humanities Council (2005) of the Royal Netherlands Academy of Arts and Sciences are working on new approaches to evaluating research in the context of broader societal questions. Also, ZonMw and the National Health Research Council (RGO) are working together on a system for the evaluation of societal quality of health research (RGO 2007).

29   Letter of 7 November 2005 regarding advice nr. 62 of the Advisory Council for Science and Technology (AWT) in which the minister of Education and Sciences declares that she will bring to the attention of the boards of universities various methods such as the sci_Quest method and that she will suggest incorporating such methods into the regular evaluations according to the Standard Evaluation Protocol (SEP), which serves as the national system of evaluation.

30   Various faculties in Dutch universities have conducted self evaluations in recent years in which the societal impact of research was included. Also, in the sector of professional education we see a rising interest, in particular since research has become part of their task description. The Hogeschool Utrecht is currently involved in a pilot study to develop a method based on the sci_Quest method in order to evaluate the research of their lectors (paragraph 4 of this introduction).

may vary from industry and governmental institutions to international organisations, NGO's and the general public. Gibbons (1999) refers to this new relationship between science and society as a 'new social contract' which not only entails the participation of a variety of actors in the knowledge production process, but also has consequences for the ways in which knowledge communication takes place and even for the kind of knowledge that is produced. Communication of knowledge under the new social contract is no longer uni-directional (from (academic) science to society), but can instead be described as a circulation process in which experts from various backgrounds and in various societal positions exchange and produce knowledge. The knowledge that results from such an iterative and broad interaction process is referred to as 'socially robust' (Gibbons, Nowotny and Scott, 2001, chapter 11). In addition to the 'mode 2'concept for research introduced in their famous 1994 book, their follow-up book introduces the 'mode 2'society, and focuses on the interaction between the two. Mode 2 society is characterized by increasingly fuzzy boundaries between state, market, culture and science, and the same goes for the boundaries between universities, research councils, government establishments, industrial R&D and other knowledge institutions such as public statistical bureaus and private research agencies. The 'mode 2' society plays an important role in the production of knowledge because knowledge is subjected in this interactive process to frequent testing, feedback and improvement from these different constituencies. (For a further elaboration on this, see chapter 3, section 3).

Areas for which such broader interaction is important are likely those with more or less established links to some kind of societal practice. While evaluation mechanisms have developed over time for areas with a discipline-oriented practice (most notably peer review and bibliometric methods such as publication counts and impact scores), this is not the case for areas with a broader orientation. And yet, we are talking here about a large group of research areas: a considerable part of the broad social sciences, the humanities and areas that are almost by nature inter- or transdisciplinary, such as agricultural sciences, environmental and climate studies, health sciences, technical sciences, development studies, etc. Researchers in all those areas feel the limitations of evaluation methods that judge their research by only (or primarily) looking at output in the form of publications.[31] And even if there is room for incorporating other aspects in the evaluation, the problem is that, unlike in the case of bibliometric methods, there is a lack of consensus about what methods to use, and/or a lack of broadly accepted methods. The Dutch national evaluation system pays attention to this problem, but it offers no solution.[32]

Over the past decade, we have seen a growing interest in these questions regarding the newly developing relations between science and society, the changing knowledge production and the need for more comprehensive forms of evaluation to assess multi-, inter- and transdisciplinary forms

---

[31] See for example Judging Research on its Merits, p. 17 and further for a discussion about worries in the ESF Standing Committee for the Humanities about the lack of adequate indicators for evaluating the humanities. See also Fischer et al. 2000 and the Allen Consulting Group 2005.

[32] In the Standard Evaluation Protocol 2003-2009 for Public Research Organisations, p. 10, relevance to the societal context is one of the four main criteria, but it is not elaborated in any substantive way.

of research.[33] Despite this, there is a lack of progress on the subject of comprehensive research evaluation. An important cause might be that, even though there is a fairly large international community of scholars who attend to evaluation in general, the number of scholarly journals that address the issue of research evaluation is relatively small[34]. Even then, the topic is so broad that it is hard to get an overview of what is being published, in particular because much is published in the gray literature. So while the practice of evaluation is compared and analyzed in several publications,[35] Shapira and Kuhlmann (2003) note that the field abounds with methods and techniques but that "nothing approaching a dominant institution or methodology exists." (p.20) Perhaps the most important reason for the lack of agreement is the fact that a more comprehensive kind of evaluation demands knowledge of the heterogeneous context of research, basically society at large. Arguably, this context is as dynamic as science itself. And while the assessment of 'scientific' quality mainly deals with the international scientific community, the assessment of societal quality deals – in principle – with the whole society. Research in agriculture, health, environment and many other fields might affect millions of people, and the questions that new research and technologies raise are debated in many sectors of society.

Therefore, we agree with the editor of the *British Medical Journal* that measuring the social impact of research is "difficult but necessary." (BMJ, vol. 323, 8 September 2001, p. 528) Moreover, he points out the remarkable dilemma that much research which is considered of the highest scientific quality has no measurable impact on health, while work that probably would not be considered of highest scientific quality (he mentions cost effectiveness studies) "may have immediate and important social benefits." (*ibid.)*

What this shows, and it is another cause of the difficulties in finding agreement on an approach for the evaluation of societal quality, is that the relation between research and society is indeed a complex one. This is even more the case now than when Daniel Bell commented in his 1973 book on the coming of the post-industrial society. According to him the biggest source of tension for science in such a society was "the relationship between the 'charismatic community' that allocates recognition and status (through the peers) and the bureaucratic institutions that allocate money and facilities."[36] To simplify a bit, when research has to deal with judgments of the scientific community, scientific quality is the most important issue. When research has to deal with society, however, a wide range of issues from ethics and safety to economics, legal issues and politics also come into play. The potential need to give all these issues a place in the evaluation process is perhaps the main reason why until now it has proven such a difficult topic. Bell essentially referred to the

---

33   There's an abundance of (gray) literature and conferences in the last decade. See for an extensive overview www.eric-project.nl.
34   There are many journals on evaluation in general, such as *Evaluation and Evaluation Practic*e, and some that focus on research such as *Research Evaluation.* Some other journals publish regularly in this area, such as *Research Policy, Scientometrics* (mainly on bibliometric evaluation), *Science, Technology and Human Values, Science and Public Policy*. Furthermore, organisations like the OECD and many other (research) organisations involved in science and technology policy (such as SPRU at Sussex University) publish regularly on the topic.
35   For example: Geuna and Martin 2001.
36   Bell 1999 (1973), pp. 385-386.

professional communities within science (institutes, associations, grant-giving organizations), but since then much more of the decision-making power with regard to scientific development became constituted in other spheres – political, economical and lay-organizations. One of the consequences has been that the agenda for science receives a wide set of social influences, of different weight obviously, but nevertheless bringing demands and limitations for scientific research to the table.

To be brief, over the last half century or so, we have seen a growing complexity in science, society and evaluation, characterized by a blurring of the borders between the three. In science, mono-disciplinary research that is primarily oriented toward the scientific community gave way to multi-inter- and transdisciplinary research. Typically, MIT research is oriented much more to societal problems than is mono-disciplinary research. Examples of the intertwining of research and society abound, but what comes almost immediately to mind is the development of the internet which increasingly intervenes in virtually every aspect of daily life, while internet users make increasingly demands that require an almost instant updating of the technological capabilities. For Manuel Castells, in his seminal trilogy about the network society, the internet is merely a metaphor for the broader development of society that he describes in terms of a process in which the major social spheres (politics, industry, science, but also crime) connect thanks to the digital technology with the financial markets as focal points (Castells, 1997). In his work, science and technology (he focuses mainly on the latter) are increasingly part of a global network economy in which their role is subject to the fast pace of transient network constellations. He paints a dynamic, but inherently uncertain picture in which not only the boundaries between the different social spheres are blurring, but the social spheres themselves are constantly changing, as are the power relations between them. And yet, they are intimately connected because the distance between the various spheres is practically zero thanks to the internet.

This is not the place to go into Castells' immense research project about the network society; suffice to say that research and technology are part of an intricate societal network that is developing with great pace and that evaluation, one way or another, needs to recognize these complexities. To a certain extent, evaluation has gone through various changes trying to meet these developments, roughly speaking, from an inward-oriented tradition of peer review with a focus on past performance to an external-oriented and forward looking practice. In these newer forms of evaluation, stakeholders play a more or less significant role, and their presence in the evaluation process distinguishes it from previous generations of evaluation. Typically, these newer forms are more oriented toward improving than judging, and they move away from a clear separation of evaluators and those to be evaluated.[37] However, we are far from a consensus about how to account for the demands and interests of these stakeholders when assessing the innovative potential (both socio-economically speaking and in terms of the technological possibilities) of research.

---

37   Gubba and Lincoln (1989) coined the stakeholder oriented evaluation 4th generation evaluation. They aim at moving beyond the previous three (characterized with the key-words measurement, description and judgment) which in their eyes are claiming to be objective and or scientific in their approach, while evaluation is fundamentally a social, political and value-oriented activity (see also paragraph 3.3 of this book).

Lack of agreement regarding how to judge the societal relevance or impact of research and consequently how to weigh that relevance or impact in evaluation procedures and science policy more generally, is a problem that is becoming more and more pressing. We have seen over the past two years or so that this problem receives attention in more and more places, policy discussions, conferences and the literature.[38] As mentioned above, the recently initiated ERiC-project aims to chart all these developments on a single map, as it were, and to stimulate a breakthrough, preferably also on a European or wider international level. Interest in this topic exists globally, in countries such as Canada, Australia and a number of African countries.[39] But initiatives like ERiC that focus on the evaluation of research in its broad context can only be successful if they also include that context in their analysis. The next step, then, is to consider the impact that this will have on science (or research) policy, both at the national and the local level.

1.3   Evaluation, policy and learning

It is clear from the preceding paragraphs that both the area of the sciences and their societal position have shown significant changes. The question we are concerned with here is what the consequences of these changes are for science policy, and more specifically, what the consequences are for the role and function of *evaluation processes* as part of the portfolio of instruments used to reconsider the functioning of research. Presently, these questions are usually posed in the light of European and global agendas (Lisbon, Millennium goals), with an accent on the innovative potential of research. It would appear that these consequences are considerable.

Mode 2 science

A method to evaluate 'science in context' has to face two challenges: first, the changing position of science and second, the changes in its context. As we mentioned above, the transformations we are discussing have been described in terms of a transition from 'science' to 'research' (Latour 1998) or, in the terms of Gibbons *et al.* (1994) as a shift from 'mode 1' to 'mode 2' science. According to Nowotny, that shift has the following characteristics (Nowotny 2003). First, mode 2 science is known for its strong focus on the context of application, which is not a synonym for 'applied science'. Apart from the fact that the distinction between 'fundamental research' and 'applied research' is quite arbitrary, it indicates that this kind of science is involved with what Latour has coined 'matters of concern', topics and problems that come into being outside the laboratories in a societal context (Latour 2005). Issues such as GM Food safety, privacy and ICT, environmental

---

38   Both KNAW and AWT (working program 2006) issued a number of publications on this subject, such as *Judging research on its merits* (2005, RGW and SWR), *The societal impact of applied health research* (RMW, 2002), AWT-advice nr. 62 "De Waarde van weten" (The value of knowledge) and the report Kennis voor de samenleving (brugfunctie TNO en GTI's, vraaggestuurd onderzoek) (Knowledge for society, a bridge function of TNO and GTIs, demand oriented research).

39   Thirteen African academies of science, together forming the NASAC, on a joint conference in Amsterdam in 2006, showed great interest in the method presented in this book. Clearly, research in development countries is for a large part unthinkable without proper involvement of stakeholders.

problems and new infectious diseases, are not only 'scientific problems'. They are political and societal problems as well. Second, mode 2 is characterized by MIT cooperation. These so-called 'matters of concern' require an approach in which various disciplines and sorts of expertise collaborate. Third, it emphasizes the crossover between institutional boundaries. Boundaries between academic, governmental and industrial research centers tend to become less relevant when it comes to the execution of 'strategic research' or 'socially relevant research'. Fourth, it pinpoints to an extension of the number and the kind of actors that are engaged in the production of research. Core-members of the research teams may very well be not only academic scientists but also professionals, lay-experts and experienced citizens. Fifth, and most important for this study, mode 2 science is accompanied by the development of new forms of quality assurance.

At the very least, these new forms of quality assurance extend the methods that we know so well in 'mode 1' science. Firmly grounded in the traditions of the various academic disciplines, quality assurance was predominantly a question of 'peer-review'. The 'colleague' and the 'connoisseur' are the archetypical actors in that tradition. In 'mode 2' science however, the question of who the relevant 'peers' are to judge the quality of science is less easy to answer. The circle has to be broadened to an 'extended peer review', including relevant stakeholders, strategic advisors, governmental delegates and professionals.

Although some authors (see note 21) question the idea of a radical shift in the functioning and the positioning of the sciences, many other analysts of science and innovation underline these tendencies. Funtowicz and Ravetz (1990), for instance, label these changes as the development towards a 'post-normal' science. With that term they indicate that the questions at stake in complex issues demand treatment 'beyond' the normal scientific methods. Complex issues require a normative and political treatment as well. Etzkowitz and Leydesdorff have proposed modelling innovation as a complex system *indicated* by a Triple Helix of University-Industry-Government relations (Etzkowitz and Leydesdorff, 1995). From our perspective the relevant question here is how these changes influence the way we talk about the 'quality' of science and how we should evaluate this. When science is more and more intertwined with its societal context, how is it possible to make judgments about the social organization of science and its financing structure, about the economic and societal benefits science produces or about the value of new ideas, new theories or innovations? The consequence of these changes is that we need to talk about science, research and knowledge in a new way. Instead of finding knowledge being judged in well bordered domains by well known procedures, we are confronted with a knowledge production system that has to deliver reliable and 'socially robust' results.

The shift to governance

The coming-into-being of a mode 2 science is not a 'stand alone', to borrow a phrase that is generally used for a PC not connected to a network. Society itself has become much more complex because the borders between different social spheres (public/private, state/market) have become

increasingly fuzzy. At the same time, a key-characteristic of the growing complexity of society is what Habermas (1985) called *Die Neue Unübersichtlichkeit*: a far-going social fragmentation that makes it impossible (and even undesirable) to steer society as a 'whole'. As Castells has put it, there is no more room for a 'cockpit' that can function as a societal power center.[40] Nowotny *et al.* (2001) label this change as a shift towards a 'mode 2 society'. The boundaries between the different institutional domains (science – society, state – market) become blurred. The existing institutions are to a lesser and lesser extent able to represent the actual social problems and face the danger of losing connection with what goes on in society. As a result, the capacity of governments to design, execute and implement policy-programs successfully, diminishes.

This wider perspective on the societal context of mode 2 science is of interest for the topic of this book because it makes clear that, for re-thinking the issue of evaluating science, evaluation processes have to take this new social complexity into account. At this point, the new questions that arise are not unique for the area of science policy. Over the last decades, all policy-areas have, in some way or another, been facing this complexity and have been forced to develop new policy-instruments to steer, execute, implement and evaluate policy-programs.

The changes in society that call for these reinventions of policy were stimulated by very specific measures of the administrations in Western nations after the mid-1980's. Prime examples are the liberalization of former state-dominated public services (public infrastructure, railways, communication-networks, energy-services) and the decentralization of government tasks. As a consequence, governments had to redesign their portfolio of policy-techniques to manage their new relationship with these liberalized agents. This swing in the polity was characterized, among other things, by the following tendencies: a shift from a central rule approach to decentralization and liberalization; a shift from hierarchy and bureaucracy to more flexibility and market mechanisms; and a shift from a supply-driven system to a demand-driven system.

The attempts to find new ways of policy-making to cope with the changes in society, have been gathered together under the concept of 'governance'. Governance is a term that encompasses many meanings and because of its growing popularity in political theory, its ambiguous application has only become worse. According to Mayntz (1998) the word 'governance' for a long time simply meant 'governing', referring to the process aspect of government. Today, however, the term governance is most often used to indicate a new mode of governing, meaning a more co-operative form, different from the old hierarchical model in which state authorities exerted sovereign control over the groups and citizens making up civil society.

Changes in policymaking: new mechanisms for learning

The question of how to deal with the fragmentation of governmental agency on the one hand and the proliferation of knowledge on the other, is not exclusive to science policy. In almost all areas of government policies we can observe the search for new policy techniques that take account of the

---

40   Manuel Castells during a debate in De Balie in Amsterdam, 21 April 2002.

growing complexity of modern societies. A closer look at these techniques might be of great help for the further development of evaluation tools for science and science-policy. These governance-techniques, as we can call them, share an attention for a multi-level, multi-actor, problem-oriented and decentralized way of policy-making that resembles the demands we are facing while considering the policy-making processes needed for the position of the sciences in its mode 2. The point of departure for this search for new techniques is the shift from the old command & control steering mechanisms to more flexible techniques. The first wave of governmental reform-programs initiated in reaction to this growing complexity started in the 1980s. It led to a whole new policy school, encouraged by authors such as Osborne and Gabler (*Reinventing Government*, 1992). The dominant paradigm to arise from this shift, especially in the Anglo-Saxon countries, was the so-called New Public Management Program (NPM). Its characteristics are a strong division between political decision-making and policy-execution, a separation of responsibilities, an emphasis on mechanisms for audit and control, a decline of the autonomy of professionals and the definition of citizens as clients, users or consumers of public services.

This New Public Management worked quite well in sectors where the costs for public services were rising out of hand and where professional alliances formed a blockage for political transformations and policy change. Initially, it was welcomed with great enthusiasm in the Netherlands as well as elsewhere. But because of its focus on *distrust*, based on the underlying Principal-Agent concept, it does have certain disadvantages (WRR 2004). First, it creates a lack of communication and lack of shared knowledge and information between organizations. Second, it establishes the dominance of management over professionals. Third, it leads to an enormous growth of controlling bodies for audit and inspections (Power 1997 speaks about the coming-into-being of an *Audit Society*). Fourth, it encourages the creation of pseudo-markets and, as a consequence, pseudo-clients and pseudo-competition. Fifth, it shapes a normative and political void: who's responsible, who's accountable?

Where the NPM program focused on a clear division of responsibilities and a restoration of the 'primacy of politics' other governance programs have emphasized the role of co-operation and co-ordination processes to deal with the new social complexities. This co-operative form can be fulfilled in many ways, with different implications for the role of the state and the formulation of 'common goods'. Schmitter (2001) stresses the mutual interdependency between the multiple actors involved in governance. He describes governance as a method or mechanism for dealing with a broad range of problems and conflicts in which actors regularly arrive at mutual satisfactory and binding decisions by negotiating and deliberating with each other and co-operating in the implementation of these decisions. However, in these arrangements, governments 'ain't in it for nothing', and still hold the final authority in many cases. In general, governance is understood as a mechanism for co-ordination by the state, involving private actors (NGO's, firms and lobby-groups) in the process of policy-making and policy-implementation.

Most of the authors involved in the business of 'governance' regard it not only as a tendency but as a necessity as well. A general diagnosis is that to deal with the problems of technological societies, traditional reflexes of the state (for instance by 'command and control') are less and less adequate. Pierre and Guy Peters (2000) therefore explicitly mention the increasing complexity of modern society as a trigger for the coming-into-being of so called 'governance' arrangements. These complexities require new sources of expertise, which make governments more dependent on external sources of knowledge.

Experimentalism

One of the most recent trends in governance is to broaden this program of co-operation and co-ordination processes with techniques that share a special attention for learning mechanisms. Inspiration for the creation of these techniques is to be found in recent developments in political theory that show a revival of attention for *pragmatic* modes of government (Cohen and Sabel 1997; Sabel 2004; Dijstelbloem and Meurs 2007). In this line of thinking society is not compared with the distant ideal of a free-market situation, but is viewed as a *social laboratory* in which science and government share the task of identifying new public problems through a process of *inquiry*. To give body to this process, several techniques can be discerned that share a focus on learning mechanisms. Key-words are the focus on more deliberative decision-making processes, problem-oriented steering mechanisms, co-operation between state and non-state actors, functionally organized institutional arrangements and feedback processes. Taken together, these learning-based policy techniques are to be considered as a coherent program for the 'governance' of those public or semi-public sectors that function at 'arm length' from the state. Science clearly is one of those sectors. Especially mode 2 science shows all the institutional characteristics of a policy domain where governance structures might be of help. Mode 2 science is a strongly decentralized, cross-institutional and multi-actor practice.

For an evaluation of 'science in context' these learning techniques deserve strong attention. The key idea of this pragmatic way of thinking is its strong focus on experiments. By experiment we don't mean that 'society' has to be managed in a scientific way (as in a technocracy) nor that policy programs are to be developed in the Popperian way of piecemeal social engineering (small steps, trial and error). In the pragmatic way of thinking, experimentalism means that new problems or changes in society (such as the crystallization of mode 2 science) ask for new procedures and institutions as well. The question of how to deal with new problems and of whether our current ways of thinking and organizing are still valid, are part of a joint program. As a process of inquiry, policy processes in general and evaluation processes in particular are to be viewed as a twofold learning process. First, we have to focus on the new questions that arise when circumstances are changing. Second, we have to reflect on the adequacy of our administrative routines. We constantly have to experiment with the institutions at hand.

Implications for evaluation

What concerns us in this context is how to arrive at an evaluation method that aims at learning processes. How to profit from the policy techniques that have been developed in the 'governance' of other sectors that share some of the characteristics of the present situation in the sciences? The paradigm shift that is needed will have to change the focus from distrust to trust. Such a method will have to serve several demands and can be described in terms of the following characteristics:

» Evaluation is not the same as accounting and control; that is, the evaluation of output in terms of certain benchmarks and indicators. The method we propose aims to include a form of *second order learning* that also puts the meaning of the benchmarks and indicators that are used into question. It therefore stimulates not only first order but also second order learning processes by way of reflection, debate and an ongoing *iteration* between goals and methods.

» It is anti-dualistic; that is, it strives to break down the dichotomy between the method of the evaluation process and its goals. As a consequence, it stimulates discussion about the choice of indicators and the process of evaluation.

» It is pluralistic, to the extent that it will have to acknowledge the variety of research programs, the variety of ways of knowing and the variety of products of knowledge (scientific publications, patents, new instruments, social benefits).

» It is stakeholder oriented in the sense that it has to take into account the fact that academic science does not have a monopoly on knowledge production.[41]

» It pays attention not only to the input in research (people, money, apparatus), and its output (publications and other products), but also to the 'throughput' (that is, processes to mediate with the environment, for example co-operation, strategic alliances), and it implies discussions about the strategic positioning of a research program, thus giving deliberation about public goals and public methods a chance.

Taken together, these principles form a program that combines some of the lessons of classical pragmatism (notably the anti-dualism) and new governance policy-techniques (especially the mechanism for co-ordination and co-operation) that share a focus on 'learning processes'.
In the method presented in this book, we have tried to account for many of these developments, but we do not pretend to have found the final solution. What we do think is that through our empirical and theoretical work we have been able to come up with an approach that in its four main parts does justice to the comprehensive character of both present-day research and questions about its value for society. However, we see this method as the beginning of a further methodological development that will be accomplished after many more experiments. These experiments need not be limited to 'academic' research. As we show in the next section, other professional sectors may also serve as breeding grounds for tests and experiments.

---

41   In our approach the perspectives of stakeholders are given a central role both in terms of information and organization of the evaluation process (see Gubba and Linclon 1989).

1.4 Learning by experiment

An example of the emergence of new forms of knowledge production that transgress traditional disciplinary boundaries and primarily work in a context of application is the recent introduction of research as a pursuit to be carried out at the universities of applied sciences in the Netherlands (Hazelkorn, 2004).[42]

The design of a quality assurance system at one of the universities of applied sciences (Hogeschool Utrecht, HU) was linked deliberately to the more general discussion taking place in the Netherlands on the necessity to incorporate a wider variety of academic functions and activities into the assessment of research. In 2005, the HU started a pilot project in which the method presented in this book was taken as a starting point. The goal was to develop a quality assurance system for its newly developed research function that could perhaps then be widened throughout the whole system. This makes it an interesting case for discussing the relation between the sci_Quest method and developments in research and research policy in the Netherlands.

The previous lack of a research tradition in the applied universities is the result of the binary system in higher education in the Netherlands, separating research based universities from the institutions for higher professional education and applied sciences that are comparable to (former) polytechnics in France and the UK or the German 'Fachhochschule'. With the introduction of the Bologna process this strict (legislative) boundary between the institutions was abandoned in the Netherlands. Consequently, the universities of applied sciences now can offer master programs with an academic or a professional orientation.

Following the Bologna declaration and the consequential ambition of the Netherlands to be one of the leading countries in Europe with regard to education and research, a research function was introduced in the applied universities in 2003. Evidently, research was felt to be a core element of innovation in higher education and professional development, and the applied universities were expected to play a major role in this development. Research in those schools is supposed to aim at strengthening the quality of education, improvement and development of a profession, contribution to innovation and participation in societal and entrepreneurial problem solving and knowledge production.

To accomplish these objectives of the new research function, the specific structure of *lectoraten* (as a counterpart to research university professorships) was devised, consisting of a *lector* and a *kenniskring* (knowledge-group). The research task of the *lector* differs considerably from the academic scholar. His or her task is to contribute to the education of 'innovative professionals' by involvement in curriculum development, training of the teaching staff, circulation of knowledge

---

42   As we explained in the beginning of this introduction, the Netherlands has a binary system of higher education: research oriented education, offered by universities, and higher professional education and applied sciences offered by hogescholen. The official English appellation is universities of applied sciences.

and the production of new knowledge/doing research. The *kenniskring* consists of members of the teaching staff who get involved in research and development activities closely related to curriculum development and (regional) societal or entrepreneurial problem solving. In this way, the teaching staff acquires research qualifications and contributes to innovation of the curriculum. Traditionally no research funds were made available to the applied universities by the government. Since 2001 government funds were allocated by the branch organization SKO to the *lectoraten* on the basis of a set of criteria that meet the goals of the *lectoraten* (see below)*.* As of 2007, these funds are part of the lump sum of the institutions. In 2009, a national system of research evaluation and monitoring will be in function for the *lectoraten.* The research function of the universities of applied sciences is expected to develop into regional knowledge centers for societal organizations and small businesses. The overall landscape of research activities in the Netherlands has been broadened by this introduction of *lectoraten* in the umiversities of applied sciences.

Research in the context of these institutes is obviously defined differently than is academic research. It refers explicitly to its value for the broader societal context. This connects in an interesting way to the new approaches of evaluation that regard research in the context of societal questions. *Lectoraten* now take a legitimate position in this particular definition of what constitutes research and are, in a sense, a new branch on the research tree. If it is recognized that several equally valuable forms of scholarship exist – *discovery, integration, application and teaching* (Boyer, 1990) – they can possibly be incorporated in a new more comprehensive evaluation system, not only including international scientific and disciplinary reputation but also 'valorization' of research, which implicates its economic, societal and cultural value.

To further stimulate the development of *lectoraten* and the evaluation of their research function, the Hogeschool Utrecht has started working with the sci_Quest method.[43] The HU initiated the development of a quality assurance system for this newly introduced research function. The objectives of the quality assurance system include: i) concentration on strategic goals of knowledge circulation and research; ii) contribution to the "professional identity" (standards, practice and ethics) of lectors and iii) development of a framework and of middle-management instruments. Recently two other institutes joined the effort of the Hogeschool Utrecht in the sci_Quest pilot: Fontys Hogescholen and the Haagse Hogeschool.
The challenges for the development of a research function in the applied universities have to be taken into account, including the lack of a tradition in research, the deficiency of the staff's research qualifications and missing infrastructure (Hazelkorn, 2004). Initial evaluations showed that conducting research by *lectoraten* was not getting sufficient priority next to curriculum development, improvement of the professionalism of teachers, networking, consultancy, etc.

---

43   According to the Dutch Higher Education law, institutions are accountable for the quality of their education and research. That is why institutions themselves are involved in the development of new quality assurance procedures, as for instance in the case of the *lectoraten*.

The international report of the committee Abrahamsen (2005) on the degree structure in higher education in the Netherlands remarked that: "The need for applied research in relation to professional education was not widely felt in the Netherlands." However, to be related to the scientific community's discussions on the reliability of knowledge, was considered an important aspect of the professionalism of *lectors*.

Next to that, the first policy evaluation showed a picture of a diverse mix of *lectoraten*. The emphasis on each of the objectives differs between *lectors*. These differences might be legitimated in relation to the professional or scientific field and the context of application. The differences should therefore be acknowledged in the assessment procedures. In this stage of development, where the standards for the *lectoraat* are still being developed, the assessment only can have a formative character.[44]

The research function of the universities of applied sciences aims at innovation, contextual problem solving and knowledge circulation and therefore includes *mode 2 knowledge production*. The experiences of previous pilot projects that employed what we referred to as the 'sci_Quest method' to evaluate Agricultural Sciences and Pharmaceutical Sciences at two Dutch universities, were used to develop further an appropriate evaluation methodology for research at Dutch professional universities. This sci_Quest method, which encompasses both mode 1 and mode 2 research activities and acknowledges legitimate differences between various kinds of research, was found to be a valuable starting point for the development of an appropriate quality assurance system. The sci_Quest method was considered useful because it not only takes results and the recognized reputation of research and researchers into consideration. It also evaluates 'science in action', 'research in the context of application', knowledge circulation and knowledge production in several societal domains. The method takes into account that innovations develop through interactions and mutual influence of varied actors. The approach in which the self-image and mission of the research group is confronted with empirical reconstructions of the interaction with the context and the involvement of users in the innovation or problem solving process that contributes to feedback and the formulation of the future strategy, is seen as appropriate because it broadens the concept of research and strengthens the 'professionalism' of the *lectoraten*.

The HU decided to develop the pilot project entitled *sci_Quest/lect* because the method made it possible to cover the wide array of objectives involved. Specific adjustments had to be made to cope with the peculiarities of the *lectoraten*. In the first place they are small compared to the average research program in research based universities. This makes it difficult to develop robust indicators. The 'professionalism' of *lectors* still has to develop. The institutional policy will have to relate to these 'professional standards'. Cooperation with stakeholders has been and will be central to the function of the Hogeschool as a regional knowledge center. *Lectors* are expected to contribute by applied research in cooperation with these stakeholders. The crucial evaluation criterion by

---

44 Formative evaluation is typically conducted during the development of a program or product as opposed to summative evaluation that regards the results or outcome (Scriven 1991).

which contextualized knowledge production is judged is no longer simply whether it constitutes good science. Now it is subject to the question of whether it provides added value for stakeholders. The *sci_Quest/lect* method was further developed and defined in an interactive process. This contributed to the development of the 'professional standards' of the *lector*. Because the building of ´knowledge networks´ was considered to be an important task of the *lectors*, the domain of 'networking' has been introduced to the evaluation process. This domain includes cooperation and coordinating expertise for knowledge production in contexts of application. The other domains are: professional competence, education, science and public policy. Also the indicators for the embedment and performance in these domains are defined following the method and are extensively described in this book. The *lectors* valued the opportunity to develop a quality assurance system that is attuned to their specific research function and therefore relevant for the definition of professional standards and improvement. They emphasized the formative character of the assessments in this stage of development of the *lectoraten*.

The pilot project started in the spring of 2006. The ambition was to present the empirical results to an external assessment committee by the end of 2006. This was possible because most of the needed information was readily available and only had to be adjusted to the indicators for each of the domains. This was found to be a time consuming task. The problems with the projection on a grid-like graph representing the involvement in each of the domains that are discussed elsewhere in this book were not easily solved. The stakeholder analysis was considered to be of much value and very relevant for the review of the research function of *lectors*. This emphasizes the process of contextualisation that is going on in the knowledge production of the *lectoraten*. The results of the pilot project are not yet available, but will be made available on the website of the ERiC project. The project demonstrates the growing fuzziness of institutional boundaries in knowledge production as well as the recognition of the expansion of scholarship, not only restricted to disciplinary boundaries, but including teaching and contextual problem solving. Developing a quality assurance system for *lectoraten* underscored the fact that research includes a broad array of activities that have to be taken into account in a comprehensive assessment of its scientific and societal value.

1.5   Structure of the book

This book has four parts, called 'the issue', 'the approach', 'the method' and 'summary and future'.

1.5.1 The issue
In this part we elaborate on the issue of comprehensive research evaluation in relation to the societal/policy context in the Netherlands. It is descriptive and problem-posing. The main question is: What is the scientific, political and societal need for a more sophisticated and comprehensive way to evaluate research?

In our view, research quality is a concept that is both relative and comprehensive: 'relative' in the sense that it is pending on relations and context, 'comprehensive' in the sense that it encompasses the broad activities of contemporary research. The distinction between 'scientific' and 'societal' quality is to a large extent artificial and not very fruitful for the development of comprehensive evaluation systems. The central question is how to deal with the fact that a productive discussion of the quality of research requires a forum that represents the various aspects of the research (for example scientific value, health risk, ethical questions, legal issues, etc.), while traditional evaluation mechanisms are designed on the assumption that the scientific community is its own best judge. We discuss the policy background of this discussion and the need for a new approach as articulated in the various scientific and societal fields.

1.5.2 The approach

What should the main characteristics be of an approach that tries to review science in a broader perspective? To analyze that, we reflect on the main theoretical and empirical considerations. We discuss the scientific literature that we used and present its main conclusions. Based on that discussion, we develop four phases in the evaluation process that constitute the 'building blocks' for our methodological program: i) identification and/or formulation of the mission and self image; ii) construction of a research embedment and performance profile (REPP); iii) stakeholder analysis and iv) reflection and strategic debate.

1.5.3 The method

In this part we will elaborate on our method using examples from the two studies we conducted. What normative and methodological problems arise when applying this approach? We address some of the major problems we encountered and explain the decisions we made to alter the first model (used in the agricultural sciences) when we conducted the second (in the pharmaceutical Sciences).

We follow the four main parts of our approach:
1. identification and/or formulation of mission/self image;
2. construction of a research embedment and performance profile (REPP);
3. stakeholder analysis;
4. feed back/discussion about future strategy.

1.5.4 Summary and future

In the final section we summarize the work we did and look into the future of research evaluation, particularly in relation to the newly introduced Standard Evaluation Protocol. We do so by asking a number of concrete questions that together account for our overall approach in the matter of research evaluation. These questions are:

» What was our assignment?
» How did we interpret the assignment?
» What is the gist of our solution?
» Which problems did we encounter and how did we solve them?
» What are the main conclusions and what are the options for the future?

There are six appendices:
1. Example from the first study in Agricultural Science
2. Example from the second study in Pharmaceutical Science
3. Indicators for Agricultural Science
4. Indicators for Pharmaceutical Science
5. Information about ERiC
6. List of abbreviations

2　The issue

In this chapter we elaborate our view on the issue of comprehensive research evaluation. The main questions are: What is it and why do we need it? (paragraph 2.1).
In our view, research quality is a concept that is both relative and comprehensive; 'relative' as opposed to 'objective' in that it is context-dependent and 'comprehensive' in that it encompasses the broad activities of present day research (paragraph 2.2). Furthermore, we argue that the distinction between 'scientific' and 'societal' quality is largely artificial, and not very fruitful for the development of comprehensive evaluation systems. To assess research comprehensively, a forum for assessing a variety of aspects such as scientific value, health risk, ethical questions and legal issues is necessary. (paragraph 2.3). At the end of this chapter, we go into the policy background of this discussion (paragraph 2.4) and the need for a new approach as it was articulated in various scientific and societal fields (paragraph 2.5).

2.1　The need for comprehensive research evaluation

To determine whether we need comprehensive research evaluation, one first needs to explain his view on what research is, followed by what the consequences of that view are for research evaluation. To be short, we view research as a social activity that is part of a larger process of innovation. To assess research, one has to regard the whole process, analyze the interactions between the different key-players and decide how this process might affect the evaluation questions and answers.

On the one hand, conducting research in 2004 is very different than it was in 1964. On the other hand, researchers do pretty much the same kind of work now as they did in 1964. For the individual researcher in his lab or behind her computer screen, things might not have changed all that much. Doing experiments or trials, analyzing data and contemplating theories, still make up a large part of the daily practice of a scientist. But at the same time, much has changed in science. Clearly, computers and the internet have made a big difference, the impact of which on research practice can hardly be overestimated. Also, the social organization of science has changed. Monodisciplinary sciences, for example, have become the minority while more and more work is done in collaboration between various disciplines, be it multi-, inter- or transdisciplinary (MIT-research). Academic research has come down from its proverbial ivory tower (not always on a voluntary basis) and ventured into mixed arrangements with industry, government and non-governmental organizations. The relationship between fundamental research and societal application of research results has changed from a uni-directional, non-communicative one into

one that is characterized in terms of networks and stakeholders. Big has become beautiful (at least in national and European programming) and one of the consequences is that in the daily practice of research there is more attention for entrepreneurial aspects. Doing science is not quite the same as running a business, but many scientists do have the feeling that they have to sell themselves continuously. To be successful in that (selling) it is no longer enough to show a perfect publication record (although that still helps). Subsidizing organizations look more and more at the potential of research in terms of surplus value for the knowledge society. Industry, which over the years left fundamental research almost completely to public institutions (universities, TNO and the like), is looking for new collaborations with academic research, no doubt pushed by large stimulation funds in fields like genomics, computer science or nanotechnology.

Collaborative research in large networks consisting of partners from academia, industry and other societal entities is encouraged these days by several policy lines, both at the national and the European level. With the growth of such research, the question of quality and relevance again needs special attention. While in the 1970s and 80s governmental focus in evaluation was arguably on accountability, the importance of socio-economic and technical value is now increasingly stressed when discussing the relevance of research.

In all this, the relationship between government and the universities has changed from one of laissez faire (1950s and 60s), to more state interference and demands for accountability (1970s and 80s) and to more autonomy and self-regulation (1990s and 21st century). At the same time, following the international trend of the knowledge economy, society and policy demand more and more that the knowledge system contributes as much as possible to the innovative capacity of the nation and Europe (cf. The Lisbon declaration).

This general trend in the relationship between science and society is largely reflected in the evaluation literature. For a long time, funds were allocated to the academic system under the premise that internal mechanisms like peer review guaranteed success in science and the best results for society. Later, peer review came under fire for reasons ranging from criticism of various prejudices that led peers to favor some researchers over others to the inability to assess MIT research and the critique that peers have problems recognizing very innovative work.[45] These criticisms weakened the position of peer review and created room for other forms of evaluation. It also opened the way for more 'objective' analysis of the science system, in particular through bibliometric and sociometric methods. Still later, evaluative systems developed in which stakeholders were allowed to participate in the procedures. At the same time, the principles behind evaluations changed from a jury model in which research was graded by peers to a coaching model, in which strategic rather than judgmental assessments were made.

---

45   Discussions about the value of peer review are abundant in the literature over the last three decades. See for a recent example Hagan 2003.

2.2 Our assignment and what gave rise to it

The introduction of the standard evaluation protocol in 2003 can be seen as the result of a decades long battle between the Dutch government and the academic system. When the government introduced the conditional finance system in the early 1980s, it aimed at making the universities more attentive to what goes on in society. (Arguably, it was also an attempt to make the universities more efficient in economic terms, but that is not our point here). It was also the first serious attempt to couple national evaluation and research policy in the sense that it made national comparison possible and aimed at reallocation among different disciplines and universities. As it turned out, the universities were not going to let this happen and they managed over the years to weaken the system until it had no meaning left (Spaapen 1995 chapter 3). In the early1990s the government canceled the conditional finance system and a new system of disciplinary evaluations was introduced in which the VSNU and the Royal Academy) played a major role (particularly in constituting the external evaluation committees.

Evaluations still focused on the value of research for the research community, but there was explicit room for a so-called multi-aspect approach. Aspects such as societal relevance and long-term viability could now be assessed separately, but an integration of the different aspects was not considered possible. There was some attention for more strategic elements, such as mission statements and descriptions of research profiles. But it was left open how the committees would deal with these elements. Assessments were to extend over the entire university research operation and were conducted per discipline. In practice what counted in most evaluations was the so-called scientific quality of research.

The system operated for a number of years but came under attack toward the end of the 1990s for a number of reasons. Frans van Steijn, responsible for evaluation at the VSNU mentioned the following weaknesses that caused a broad loss of acceptance in the university world: the system was not able to cope with the growing number of MIT research programs; the system was not flexible enough to serve researchers who participated in different programs; consequently, researchers felt they had to produce too much unnecessary information; the consequences of the evaluations in terms of changing research policy were often not visible. This last argument related to the Dutch government's demand for more accountability vis-a-vis the developing knowledge economy.[46]
Growing criticism led to the installation of a committee that was to produce an advisory report for a new evaluation system. In 2001, the committee issued its report[47], in which suggestions were made for a new protocol. Its main recommendations were that the new protocol should focus

---

46   Frans van Steijn: "Standaard Evaluatieprotocol: ervaring en bronnen", presentation at a conference in Nijmegen for the introduction of a new research information system at the Dutch universities, 2003.

47   See Werkgroep Kwaliteitszorg 2001.

not only on scientific research quality, but also on aspects of socio-economic and technical value and of research management. It also advised replacing national comparisons with international comparison and adopting a more forward-looking orientation. Another important suggestion was that units to be evaluated should write a self-evaluation document for the evaluation committee.

The replacement of the 1994 protocol by a new one that largely followed committee recommendations happened in 2003 when the Standard Evaluation Protocol (SEP) was introduced after long negotiations between the universities, the Academy and the National Science Foundation (NWO). One important characteristic is that the development of the new protocol was left completely to the research system itself, under the condition that there would be an independent meta-evaluation committee to organize a national review of the success or failure of the SEP. The boards of the research institutions (universities and other institutes) were given full responsibility for the organization of evaluations, but they had to make clear to the minister what the consequences of the evaluations were for future research policy. One of the four main criteria was socio-economic relevance. In particular, it was stressed that evaluations should be organized in ways that would enable the assessment of MIT research. It was in these areas that sci_Quest got the opportunity to develop a method for doing a broader kind of evaluation. The first method, employed in 1998 in Wageningen (Agricultural sciences) was under the 1992 protocol, the second one in Utrecht en Groningen (Pharmaceutical sciences) was done at a time when the new protocol was in the design phase.

Our assignment entailed the development of a method that could be used in the overall evaluation system in the Netherlands. It meant composing a report that would represent the different research groups in a comprehensive way, so that a group of international peer reviewers could assess each group's broader range of activities in light of its chosen mission. In other words, the evaluation had to be *mission-oriented*. We were asked to provide systematic and comprehensive information on issues that regarded both the scientific and societal value of research. Also, our method was to encompass the interaction with the user or stakeholder environment.

For us, the assignment meant that finding a new evaluation method required us to understand the relations between a research group and its relevant context: how does it relate to its context, what is the role of the various stakeholders in that context, and how does that relate to the mission of the group?

2.3   Research in the context of stakeholders

Beginning with the example of research groups in the agricultural sciences, it is clear that they perform their work in a context of application (which does not necessarily means that all their research is applied). Research has to be scientifically sound and credible to colleague researchers. But it also has to meet the interests of a variegated group of stakeholders such as farmers, local and national government (regulations), consumers (preferences), etc. Research, in other words has to mediate between a multitude of interests and values. Research functions in different ways in different social domains, and innovations are more likely to succeed if the research meets the interests of the different social groups involved, and is open to external expertise. Research results have to be, in the words of Nowotny *et al*., 'socially robust'.[48] The make-up of the context of application depends strongly on the research strategy of the groups. In our evaluation research strategy is therefore a central concept.

Furthermore, research programs and their contexts share a dynamic relation. Initial goals may get lost if new opportunities emerge. Research on nitrate and sulfate cycles, for example, was initially only relevant for the agricultural production. Now it finds new relevance in research on global climate change and the greenhouse effect. The research strategy of a group may change accordingly; new projects may be formulated (and/or old ones reformulated) to meet criteria of the national and international organization funding climate research. The program then functions in a new environment with new standards for what is good and relevant research. A uniform yardstick as it is used in the more traditional evaluations would not do justice to the specific nature of a research strategy and the dynamic nature of it in relation to a changing environment.

If research is conducted in the context of application, as a rule there is a good deal of interaction with experts from other (sometimes non-scientific) areas. Many specialties within the field of agricultural sciences are problem-oriented (and work in the context of application). In the project we did in Wageningen, we looked at research groups operating in fields like crop- and grassland sciences, plant production systems, soil tillage, farm technology, irrigation and water engineering. A number of the groups were oriented toward development, i.e. 'third world' problems. All these research programs typically combine insights of several disciplines and technical expertise, operating in a policy-context in which interaction with a variation of users is a necessary condition to make solutions work.

In the pharmaceutical sciences also, we saw research that can be characterized as MIT research. Pharmaceutical research often features a close cooperation between university research and industry, if only for the fact that research groups need the funding that comes from industry. But its

---

48   The term 'socially robust knowledge' is explained extensively in chapter 11 of Nowotny *et al.* 2001. We will come back to it later. Here suffice it to say that knowledge is produced and tested not only by science, but also in collaboration with other experts.

broader societal relevance is also clear. Academic pharmaceutical research has been known to have close contacts with professionals (pharmacists, chemists and patients). The variegated character of this stakeholder environment of pharmaceutical research can easily be described in terms of the 'mode 2' research.

In both these areas, research not only transgresses disciplinary boundaries, but also those of professional and lay expertise. For research policy that tries to assess such MIT research, it is extremely difficult to understand and weigh all these different influences in an evaluation process. And this is aggravated because measurements (and data) for the so-called scientific quality of research seem to be abundant compared to the societal quality indicators.

2.4   Policy background of both studies

In 1998 we conducted the first of our two studies in the field of agricultural sciences at Wageningen University. Research there is organized along multi-disciplinary lines and is often performed in a context of application. In the last ten to fifteen years the direction of research in this sector has extended toward land use and the food chain and now also includes environment and health issues, as well as the socio-economic exploitation of 'green space'. Given this development, the university felt increasingly uncomfortable with the fact that they had to present their research according to the old protocol to committees that were not seen as very open to their specific approach to research production. Therefore, the university volunteered to have a number of groups to be evaluated through an experiment, i.e. our method.

In 2002 we conducted a study in two faculties of pharmaceutical science in the Netherlands.[49] Basically, the same motives were at stake here. The study again aimed at assessing the research effort of those two faculties in relation to their interaction with both scientific and other relevant communities (industry, professionals, policy, the public at large). But we now we were also asked to see whether the method we tested in Wageningen could be improved.

Still, the primary goal of the second project was to support the two faculties in conducting the self-evaluation required under the renewed evaluation system for academic research in the Netherlands (SEP 2003). Self-evaluation under this new system entails a report in which the research unit to be evaluated is presented in all its relevant activities and output. It also contains an analysis of the strengths and weaknesses, and a vision on the future of the group (program).

A second goal of our study was to see whether the method we proposed, including data-gathering and -analysis, was straightforward enough to be carried out with the help of the groups and/or

---

49   Sci-Quest/GRIP 2003.

faculties under investigation within a reasonable time and with reasonable bureaucratic effort. A third goal, finally, was to see whether the indicators we use in this method could be made stronger by a benchmarking operation.

The new protocol introduced in 2003 offered an excellent opportunity to investigate whether our method could be implemented within that new system. A number of the characteristics of the protocol seemed to parallel our own ideas about evaluation. The importance of self-evaluation, the accent on looking forward instead of back, the elimination of comparison on the national/disciplinary level (in favor of international comparison), and discussion about the research program's mission were all elements that – on paper at least – compare to our own approach of evaluation.

As mentioned above, the question of the societal relevance of research and how to evaluate this is getting more attention these days, not only in the Netherlands, but internationally too. In the UK, for example, a number of studies have been conducted in the last decade to assess the relevance of research to society (Lyall *et al*. 2004). The question is relevant for most scientific research, but perhaps more obvious in some cases than in others. As was the case in our studies, both agricultural sciences and pharmaceutical sciences are obvious and interesting cases.

2.5   Blurring borders between science and society

As we have seen, both pharmaceutical and agricultural research has many faces besides its activity in the scientific (academic) domain. Research in these fields is clearly embedded in the economics of technological innovations and industrial production. It is also involved in the production of social benefits, for instance by helping to produce safer foods or stimulating closer connections between professional pharmacists and patient needs. Pharmaceutical research in academia is, to give another example, closely related to several research projects with governmental institutes to analyze policy-related problems or to develop future regulations on the use of drugs.

Thus, research in these areas is characterized by a close cooperation between university research and industry, but the research is also linked to other societal constituencies such as professional- and patient-organizations. It is also frequently subject to political debate. This variegated character of the stakeholder environment of research is characteristic for much of today's research. Theoretically, this changing character of research has been described in various ways. After the so-called 'relativistic turn' in the sociology of science, perspectives on the scientific enterprise underwent rapid yet fundamental change exemplified by the emergence of new forms of research organization (Wilts 2000). The change has been alternately described as the advent of 'post-normal science', characterized by a new and value-sensitive methodology (Funtowicz and Ravetz, 1990),

as 'the new production of knowledge', or 'mode 2' science, replacing conventional forms of research organization (Gibbons *et al*. 1994), as the emergence of a 'triple helix' of intricate relations between university, industry and government (Etzkowitz *et al*. 2000) or techno-economic networks (Callon *et al*. 1992). These notions differ in terms of their particular conceptual definitions and assumptions, but share an orientation toward innovative and application-oriented research at the interface between scientific, economic and political domains.

The debates about what is going on within research communities and between them and the wider environment (the polity, society at large) regarding the relevance of research and how to assess that have been and still are rather controversial overall, in particular when the social direction and/or the demand orientation of research is discussed. The focus of the larger debate has shifted over the years, arguably in connection with changing views on the relationship between science and society. During and after WWII, it seemed self-evident that research ('operations research') was useful and served policy (military) goals. But the strong image presented by C.P. Snow in his 'Two Cultures' (1959) which opposed (natural) science to culture/society, led to a broad awareness about the different goals, interests and expectations of the two social spheres. For a while, science was able to present itself as being most effective when left untouched by policy intervention (1960s, early seventies).

Later, when universities turned into mass institutions and budgetary matters became more important, questions about the relationship were drawn into a political-administrative sphere (e.g. the 'quid pro quo' debate of what research can do for policy/what policy can do for research; Blume 1986). This marked the beginning of a stronger urge from government for research to be 'useful', and the idea that researchers should account for that in evaluations. Still later, interaction between the two spheres grew more intense, and expectations that research be useful for society increased similarly. More recently, the literature shows a growing integration of science, government, and societal actors, such as industry and NGO's, partly as a result of specific government policies (Wilts 2000). As a consequence, we see a growing interest in more specific user/stakeholder oriented questions in evaluation debates both in the literature and in national and international policy circles (Shinn *et al*. 1997, Spaapen and Wamelink 1999, Den Hertog 1996 (a report for the EU), Proceedings of the International Transdisciplinarity 2000 Conference).

To a large extent, this complicates the question of how to evaluate research. Research seems to transgress not only disciplinary boundaries (multi-, inter- and transdisciplinary research), but also the boundaries of other participants of the innovation process. The research agenda is set not solely by academics, but to a certain extent also by other stakeholders, and this is expected to be reflected in the profile of a research group or program.

The diverse environment of pharmaceutical or agricultural research, the widespread spin-off of results, their use by a plurality of social actors and the close co-operation between academic scientists and professionals from other societal domains hardly comes to light when applying standard evaluation procedures. This may lead to an underestimation of academic research groups that are closely involved in the societal domain or in industrial research – at least, their research 'profiles' tend to be distorted.

As a consequence, evaluations of research in those sciences that follow more traditional methods (primarily focusing on publications in scientific journals), often do not do justice to the broad spectrum of activities and relations of pharmaceutical researchers. Previous evaluation systems were criticized for their lack of recognition of the broader impact some sciences have. The new Standard Evaluation Protocol however does recognize this and leaves room for the development of additional methods to evaluate research in a broader perspective. We have worked out such a method.

2.6   Upshot for a method

We explain our method in detail in chapter 3; here we describe some of its main characteristics. Two premises are essential for our method:

1.   We see research, and therefore research evaluation, as a comprehensive endeavor. Research is part of a broader process of innovation in which there is no straight line from fundamental research to practical application, but an interactive and iterative pattern of mutual influencing between the different actors ('stakeholders') in that innovation process;
2.   Evaluation is mission oriented, and therefore reflective and forward looking more than judgmental and backward looking.

Essential in our approach is the involvement of all participants in the research process, inside as well as outside the research group; that is, both the researchers under investigation and the broader environment of the research groups. The environment of the group is addressed qualitatively in the stakeholder analysis and quantitatively in the REPP. The research groups under evaluation, the administrators and others in the faculty help in gathering the data (for which no standard procedures are yet developed). This way, we gain 'in-field' knowledge of a particular area that helps us determine what we need to develop as criteria, indicators, and benchmarks. Participants in the process of data gathering not only help us with the data, they also give us feed back at different times during our study. In other words, the groups and their environment become part of the evaluation process, part of the method.

A significant characteristic of any evaluation, but certainly of our method, is self-evaluation. For us, self-evaluation is more about self-reflection and future policy than about judging the past and giving a 'good' or 'bad' verdict. In that respect, we support the red line in the new Standard Protocol that stresses looking forward rather than back. This, of course, is not to deny the value of looking at past performance, which is certainly part of our method, but it is to stress the fact that this kind of evaluation is of a different kind than more traditional ones. If we characterize traditional evaluation as a 'jury' or 'verdict' system, our type of evaluation is that of a 'coaching' system. It therefore encompasses not only the so-called scientific performance, but also other aspects of the research and innovation process.

3 The approach

In this chapter we pose the question of what the main characteristics of an approach that tries to review science in a broader perspective should be (paragraph 3.1). The need for new evaluation mechanisms is on the rise especially since MIT research is growing. As a rule, most traditional evaluations focus on the so-called scientific quality of research, and most reward systems in science take the same viewpoint. From the point of view of researchers involved in MIT research this is not very satisfying because the work they do and the output they produce is – at least partly – of a different nature and does not satisfy the criteria of these traditional evaluations. Also, from the point of public interest there is dissatisfaction with the limited range of traditional evaluations (paragraph 3.2).

To develop an evaluation method for MIT research, one needs to understand what such research means in terms of the wider production process of knowledge. We discuss the current state of the relevant scientific literature and present its main conclusions (paragraph 3.3). Based on this review, we introduce the 'building blocks' of our methodological program; that is, we identify four phases in the evaluation process: identification of the mission; construction of the REPP; stakeholder analysis; reflection and comparative feedback and strategic debate (paragraph 3.4).

3.1 Strategic and comprehensive nature of MIT research

Innovation, the wider process of which research is part, defines the way we approach evaluation. It is a compound process in which many different stakeholders operate. Agricultural and pharmaceutical sciences are two good examples of scientific fields where, researchers are actively enmeshed in a complex context of stakeholders, each of whom tries to influence the research process by asserting various demands and interests. This leads to tensions among different norms and values in the various stakeholder communities. Academic researchers in general want and need to be independent, patients want safe and cheap drugs, industry wants to make a profit and the polity wants rules and regulations. Research groups deal with these tensions in various ways and form different coalitions with different stakeholders. Further, different demands and expectations develop in every subcontext, bringing different evaluative norms and values with them. Before we go into the consequences for evaluation, we need to describe the research process as we see it somewhat more.

Our method is essentially for research that is multi-, inter- or transdisciplinary. We refer to it as MIT research because, though analytical distinctions exist between the three terms, most people use the words as near synonyms. We realize that terms like multi-, inter-, or transdisciplinarity cannot

serve as discrete categories in the sense that they are mutually exclusive. Research is too complex an enterprise to be put into boxes and labeled (although thinking along disciplinary lines tends to do exactly that). Researchers do research, sometimes alone in a laboratory, sometimes in groups with other researchers, and sometimes in broader contexts, collaborating with industry, policy or others. When we talk about multi-, inter-, or transdisciplinarity, therefore, we refer to research in the first place as a social process of knowledge production. As such, knowledge and information moving between and across disciplines encounter resistance produced by disciplinary cultures (languages, methods, self-images). This hinders the communication and integration of different kinds of 'knowledge' and expertise; the extent to which this happens is a key factor in the success or failure of the research enterprise.

With this said, there are definitional and practical differences among multi-, inter- and transdisciplinarity that are worth attending to. The term transdisciplinarity emerged in the beginning of the 1990s, in particular through a book by Gibbons *et al*. called *The new production of knowledge* (1994). The authors distinguish between a *mode 1* and a *mode 2* knowledge production, the former referring to disciplinary academic work, the latter to research that takes place in a context of application, which they call transdisciplinary research. Since then the topic has been subject to debate in the international literature and at a number of international conferences that have been held on the subject (Proceedings 2001). Particularly in some scientific fields (education, health, agriculture, environmental sciences, engineering, urban and landscape studies, development studies), transdisciplinarity has become the dominant mode in which research progresses. Many other areas in society are soliciting transdisciplinary research (climate and water studies, biodiversity, transport and engineering, energy, socio-economic questions). One important difference between transdisciplinarity and multi- and interdisciplinarity is the extent of integration reached at the program level. In the case of multi- and interdisciplinary research, integration between different kinds of disciplinary expertise is often not very strong. Researchers pool their specific expertise to work on a particular topic. Multidisciplinary work is often inspired by some sort of societal question; in the case of interdisciplinarity, the work is mostly inspired by scientific developments. In either case, coalitions are loose and not integrative, although new (sub)fields might emerge out of interdisciplinary research, a recent example being bio-geology (see Thompson Klein 1990 in particular part II, for a detailed description of the obstacles in these integrative processes).

In multi/interdisciplinary research the interests sometimes overlap, sometimes not. For each stakeholder, the development of his/her own field is more important than the joint problem solving; when working on a project there is always a potential conflict between different ideas, values and approaches. In transdisciplinary research participants leave their own discipline, expertise and area behind. The coalition, though temporary, transgresses disciplines, and is therefore integrative.

Transdisciplinary research projects are conducted in collaboration with stakeholders from other societal backgrounds and expertise.

Research groups in MIT research work in a context of application. This means that research has to meet the interests of a variegated group of stakeholders that includes doctors, pharmacists, chemists, patients, local and national government (regulations). This has given rise to questions about the independence of academic research. In the international scientific field the two farewell articles of the editor of *The New England Journal of Medicine*, M. Angel (2000) on the privileged position of the pharmaceutical industry gained a lot of attention. In *The Lancet*, Wheaterall (2000) condemned the too close relationship between academia and industry, describing them as 'increasingly uneasy bedfellows'. Djulbegovic *et al*. (2000) made a similar point, focusing on how the uncertainty principle is trespassed in clinical trials (co-)financed by industry. In the Netherlands, the professional magazine *Geneesmiddelenbulletin* (2001) showed the reaction of the pharmaceutical industry and scientists. The public debate was further stimulated by De Vries (2001), followed by a discussion forum, as well as a report by the Socialist Party (Kant *et al*. 2001).

At the same time, academic pharmaceutical research is known for its ties with patient movements. Professional alliances with these groups are a further stimulation for the development of so called lay expertise and their 'cognitive mobilization' (e.g. Epstein 1996; Callon and Rabeharisoa 1998; Dijstelbloem 2000). Moreover, several research groups known as 'social pharmacy' explicitly encourage public debate, public participation and public education on the use of medicines and the pro's and con's of different treatments and therapies, closely co-operating with professional pharmacists by providing them with information and tools to collect this information (databases). Next to that, these academic groups periodically function as pressure groups by lobbying the Ministry of Health to promote further research on some categories of so-called 'orphan drugs'; that is, drugs against rare diseases that are of little commercial interest and not produced by the industry.

Thus, academic pharmaceutical research clearly has many faces besides its activity in the scientific domain. On the one hand it is embedded in the economics of technological innovations and industrial production; on the other hand it is involved in the production of social benefits, for instance by stimulating closer connections between professional pharmacists and patient needs. Furthermore, academic pharmaceutical science is closely related to several research projects with governmental institutes to analyze policy-related problems or to develop future regulations on the use of drugs. The make-up of this context of application depends strongly on the research strategy of the groups. A uniform yardstick as it is used in the more traditional evaluations cannot do justice to the specific nature of a research group's strategy and its dynamic relation with its environment.

The assessment of such research calls for a different approach than the traditional peer review and/or bibliometric analyses used in most current evaluations in the academic system.

3.2   Evaluation and MIT research

To assess research that takes place within the broader context in which interactions among a variety of different actors determine the direction of research and innovation, one needs to analyze that context. The context can be seen as an interaction pattern between different actors engaging in various cooperative efforts, not only between scientists, but also between them and other experts of participant interest groups. These interaction patterns lead to variegated forms of knowledge production that are rather different from the traditional disciplinary manifestations of research. Whether this is truly a new phenomenon, as Gibbons *et al*. (1994) claim, is matter of debate (Shinn 1999, Proceedings 2000). Michael Gibbons and Helga Nowotny characterize transdisciplinary research as follows[50]:

» Transdisciplinary research takes place in a 'context of application'. That is, different societal interests ('stakeholders') are represented in the research process. This does not necessarily mean that it is about applied research only.

» Transdisciplinary research groups are characterized by a flat hierarchical structure and are heterogeneous, contrary to (traditional) university groups.

» Transdisciplinary research does not respect disciplinary boundaries, it aims at being a forum for alternative intellectual challenges, a different research characteristic emerges. Of vital interest is that research is seen as transgressive, as something that by nature cannot be tamed by disciplinary boundaries.

» Transdisciplinary research fits in a world in which research is more and more the result of interaction between 'stakeholders'. This goes for politics, the economy, and science and technology, as for the interaction between these terrains.

» Transdisciplinary research aims at being accountable to a society that asks: "what are you doing for us?". Such does not happen in a linear relationship (researchers do research and this is translated to society), but in an iterative process in a network of interactions, in which knowledge and expertise are socially distributed. That is, the transfer of knowledge and expertise goes back and forward between different actors. The knowledge that emerges in these networks is not only scientifically reliable, but also 'socially robust'. The latter term refers to a kind of knowledge that is, contrary to what disciplinary knowledge tries to accomplish, open-ended, relative to a context, and liable to testing and validation by a variety of stakeholders.

---

50   Gibbons and Nowotny (2001), p. 67-80.

Gibbons and Nowotny add that transdisciplinary research needs new forms of evaluation and quality control. In these, the context of research should play a vital role. They stress that transdisciplinary research does not replace disciplinary research, but develops next to it, in partial combination with it. As Nowotny once put it, there are no *full-time transdisciplinarians*. Researchers, like other stakeholders in the process, form temporary coalitions to work on certain problems. Essential is the urge to work together to solve a problem. Transdisciplinary research can be defined as *joint problem solving in the context of application.*

A keyword in the analysis of knowledge production that Gibbons and Nowotny use is 'transient', meaning a dynamic field of ever changing constellations of stakeholders. These stakeholders operate in broad social and economic contexts. But when they come together to do transdisciplinary work, specific cognitive and social practices develop, in which 'accountability' is a prime value. Reflection on what is good research and how to evaluate that in the context of societal demands also comes on board. It is therefore crucial that, in our evaluation method, we focus on these transient constellations and on the interactions that take place in the knowledge production process. Finding a new evaluation method thus means understanding the relations between a research group and its relevant context: how it relates to its context, what the role of the various stakeholders in that context is, and how that relates to the mission of the group.

3.3 Heuristic

In developing our method, we built on a wide set of literature, mostly from the area of science and technology studies and the literature on evaluation. The following three areas are the most important:
1.   Discussions around the so-called 'new production of knowledge';[51]
2.   Evaluation literature, in particular the group of French researchers who developed a so-called compass card for research labs and the so-called 4th generation evaluation (Gubba and Lincoln 1989) in which stakeholders play a vital role;
3.   Work from the area of innovations, particularly studies that focus on the role of users and learning processes.

In the science and technology studies literature, new conceptual lines have come to the fore regarding the organization of knowledge production. Scholars abandon the 'old' adage that science operates in a relatively isolated social position, that it produces and validates 'reliable' knowledge, and communicates its discoveries to society through some kind of established transfer mechanism.

---

[51]   The term was coined by Gibbons *et al*. in their 1994 book in which they distinguished a 'mode 1' and a 'mode 2' knowledge type, reserving the first for disciplinary knowledge, the second for transdisciplinary knowledge.

But now, the more recent idea of science operating in a policy arena in which different actors interact (for example In 't Veld 2000) appears also to be giving way to yet another concept. In this, the dominating image of the relationship between science and society is neither that of a linear relationship, nor that of interest groups in some kind of political struggle. It is that of a more or less transient team of experts coming together from different scientific, technical and socio-economic disciplines, including users, and striving for 'joint problem solving' (Gibbons *et al*. 1994; Nowotny *et al*. 2001).

Knowledge, in this context, is seen as something essentially transgressive, meaning both that it knows no boundaries and that it can come from different directions (not only science). Connected with this transition is the changing idea of what science and doing research essentially is. The image of an ivory tower in which the truth will be discovered has been long left behind and replaced by more relativist perspectives in which science is firstly a method, not an answer to a question. Science, in this perspective, is part of a whole process through which society renews itself, that is, the innovation process broadly understood. In this process several key actors may be discerned (see Verkaik 1997), each bringing their own expertise, experience and wishes to the table. Consequently, the traditional esteem for science is diminishing as uncertainties grow – both on the side of science and of policy/society. This new situation demands different organizational approaches and different visions of science and its evaluation.

Gibbons *et al*. (1994) state that two factors are especially important in the interaction and communication among the various actors in the network around scientific groups: the mobility of scientists (because it is essential for cross-fertilization of knowledge and know how) and the way problems are selected and priorities set. How and why are some problems selected and others not, and what are the differences between fields? This connects to the third field of literature mentioned above (innovation studies), to which we will return.

We used these two factors as a heuristic in finding differences in the research contexts of groups. The context of research groups may differ considerably as some groups are mainly oriented toward the translation of practical problems into a scientific approach (*i.e.* the modeling of pharmaceutical production processes), while other groups are at the end of the application phase and engaged in designing apparatus for improved administering of tablets or crop yield.

In the more recent book, *Re-thinking science* (Nowotny *et al*. 2001), the same authors take their argument farther by focusing on the co-evolution of science and society. That is, they no longer regard the transformation of society as predominantly shaped by scientific and technical change. Rather, they portray the development of both science and society as highly interactive and, alongside mode-2 science they discern a mode-2 society. Both have become 'transgressive arenas'

in which socially robust knowledge is produced. This socially robust knowledge is a complex and easily misunderstood concept; the authors oppose it to more conventional concepts of knowledge as the outcome of science that seeks the (absolute) truth. Socially robust knowledge, the authors claim, is more flexible and open-ended, it is relative to a context, and testable and liable to validation by a variety of actors in the network. The authors stress that socially robust knowledge is not the same as socially accepted knowledge; socially robust knowledge is the product of an infiltration and improvement by social knowledge; it critically absorbs, so to speak, knowledge from all social spheres, from both science and society (*ibid*. 167).

An example of the work in agricultural research may clarify this somewhat. When we think of the work on genetically modified foods, research in that area has to attune new scientific knowledge with professional principles and societal demand. In this, socially robust knowledge is the result of an information process in which issues of scientific, ethical, social, cultural, legal, economic, etc. values are 'absorbed'.

The second field of literature on which we relied is that of evaluation studies. Here we were most impressed by two developments: work on the so-called compass card that aims to represent research groups through the variegated activities they perform in a context of application (Callon *et al*. 1992) and work on fourth generation evaluation that integrates stakeholders into the evaluation process (Gubba and Lincoln 1989).

From work on the new production of knowledge we learned that research programs develop in mutual transactions with a relevant societal environment. The success of a research program therefore partially depends on the ways in which researchers manage to connect to themes in that environment, and on the ways in which this environment absorbs ('uses') and further develops the results of the research. A similar point was already made in the area of science and technology studies a few decades ago in a slightly different context. Studies had shown that the production of knowledge in research programs is closely connected to the local organization of the research context. Researchers implicate questions and problems raised by societal actors in their programs. Such a strategy aims, among other things, at safeguarding 'resources' for research (Knorr-Cetina, 1982; Latour & Woolgar, 1979). Research programs therefore develop in different ways in accordance with a particular environment. Some programs develop primarily in connection with the international scientific community (often a disciplinary community). Others are more oriented toward European networks in which general policy questions are at stake, or collaborate with professional entities in a context of application. As a consequence, traditional qualifications for research groups such as 'applied', 'fundamental' or 'pure' do not fit the wide variety that can currently be detected. To do justice to the diversity that has developed in terms of the dynamics and organization of knowledge production, Crow en Bozeman (1987), Callon *et al*. (1992) and Joly and

Mangematin (1996) developed classifications for different laboratory profiles based on an analysis of the relations research groups had with their environment.

Callon *et al*. designed the compass card of research in which that diversity is systematically represented on the basis of empirical research. This compass card distinguishes five social domains in the context of a research program, in which separate criteria for assessment are developed. The domains are connected with particular actors and sectors. This approach helped us to develop the REPP.

Social domains or contexts for knowledge production include the international scientific community, the world of professionals, industry and the policy context. In each of these contexts, different expectations exist with respect to the research; in each, different norms, values and priorities influence the development of a research program. Interaction mechanisms and patterns are bound to differ among these contexts. Within these social domains interaction channels are distinguished that are characteristic for communication between scientists and environment and that can serve as a base for empirical research. Callon *et al.* distinguish:

» Texts, including scientific, professional *and* popular
» People, including researchers and other actors
» Artifacts, including apparatus, protocols, rules and regulations
» Money

In each of these interaction channels, a set of criteria can be defined, based on what is customary in the specific field of study.

Callon's work is geared toward developing lab profiles; it is not an evaluation tool as such. And it does not include a role for the stakeholders in the evaluation process. That is precisely what Gubba and Lincoln (1989) aim at. Their work on evaluation is rooted in the education sector, but has wider implications for other sectors where stakeholders are relevant. In their seminal 1989 account they distinguish four generations of approaches to evaluation. In the first, the focus was on measuring the performance of individuals and the evaluator was seen as a technician who applied measuring instruments. The result was a report with clear figures. The focus shifted in the second generation to the close environment of the individual, in this case the school curriculum. Object of evaluation became the mission of the school: did pupils learn what the teachers intended them to. The evaluator was expected to describe strengths and weaknesses in the school program.

In the third generation evaluation, the question shifted from "are the objectives being achieved" to "are the objectives worthwhile?" Evaluators were to judge this broader question, which had in itself not much to do with the activities in the programs at hand.

Three major flaws were identified in these first three generations of evaluation, which are of a general nature:

1. The tendency towards managerialism. Managers and evaluators decide which questions should be asked; other stakeholders are not represented.
2. Failure to accommodate value pluralism. Value pluralism is essential between different cultures within a society and, as such, constitutes a crucial matter in evaluation.
3. An over-commitment to the scientific paradigm of inquiry. Extreme dependence on scientific methods has resulted in the erasure of context, overdependence on quantitative measurement and "coerciveness of the truth". Consequently, the evaluator bears no moral responsibility for his conclusions.

To answer these flaws, fourth generation evaluation is organized by the claims, concerns and issues of stake holding audiences, and it uses constructivist methodologies. This means that the evaluation method includes all relevant stakeholders, elicits from the stakeholders their constructions of the main issues in the evaluation, and builds an evaluation context through which these different constructions can be understood and critiques taken into account.

Finally, the third field of literature we distinguished above regards innovation studies, particularly work on users or stakeholders. Since the 1970s the literature stresses the iterative character of innovation processes (Freeman: 1974; 1991; Nelson en Winter: 1977; Nelson: 1993). That is, innovation is seen as the result of many interactions between partly overlapping social networks. These networks consist of more or less stable communities of stakeholders. In the process of mutual influencing, technical, political, institutional, socio-cultural and economical (market) factors play a role. In the 'evolutionary' corner of innovation studies, technological development is seen as a process that shows parallels with evolutionary biological theory. Innovations are seen as the result of a process in which various options ('variations') are tried out in a so-called selection-environment. The search process develops along more or less stable trajectories in which rules ('technological regimes') and social structures evolved over time through interactions between different stakeholders. The process is not entirely deterministic. For example, novelties introduced in the selection environment do not, for a variety of reasons, always 'fit'. Experts might not agree, for instance, about the road forward; random elements play a role, such as unexpected developments in the field, or in neighboring fields. In evolutionary approaches to the innovation process, therefore, learning processes form an important element. In these learning processes, feedback from the selection environment contributes to the technological design. Both the learning by researchers through contacts with users, and the learning of users in interaction with researchers are analyzed. Several phases can be distinguished in the innovation process, such as an initial phase in which a societal and technological problem is articulated and, later, a phase in which the

innovation is tested experimentally (Den Hertog *et al*. 1996). In each of these phases, differences can be found between the ways in which involved actors communicate and learn.

In this context, it is expected that researchers aiming at success for particular innovations opt for a broad orientation in the environment, and develop relations with a variety of potential 'users': entrepreneurs, consumers, policy makers etc. In this, research groups are 'open systems' exchanging with their environment (Verkaik 1997). They are characterized by a continuous cycle of input, internal change, output and feedback.

The mutual exchange between the different social domains is very important for the system as a whole, its viability and its form, because that exchange is its *raison d' être*. When one wishes to assess the 'innovative power' of a research program, one inevitably has to involve collaborations and communications into the evaluation process. In science and technology policy circles, therefore, a need has developed for a method that does justice to all these factors in a balanced evaluation procedure. It is this need that is addressed by this study, which is why the evaluation method it features focuses on interaction processes between research and a relevant environment.

In the work of Den Hertog *et al.,* the importance of learning processes in the development of social and technological innovations is pivotal. They too stand in the tradition of scholars that see innovation in terms of an evolutionary process: innovation takes place in a mix of technical and non-technical networks. The authors distinguish three phases in the knowledge production, referred to as articulation, attunement, and fine-tuning. In the first phase the chances for agenda setting by stakeholders is relatively large. In the second a more or less stable environment for the particular topic starts to gain shape and influencing the direction of research becomes less easy. In the last phase, experiments or trials outside the lab take place and the possibility to influence is minimized for the stakeholders.

In the first of our two studies, we focused on learning processes and tried to identify different learning environments. Since the evidence we were able to gather was not very strong, we decided in the second study to leave this and switch to a the role of stakeholders in the research agenda setting. For example, do well organized patient groups with a considerable amount of funding have more opportunities to influence the agenda of academics than relatively small and/or 'poor' groups which may encounter many difficulties in gaining access to actual research? Or, do ties between industrial research and public science lead to new forms of screening off science? And do these ties become stronger in a situation where public funding is insufficient? How transparent is the agenda setting in this new production of knowledge? Is a gatekeeper still drawing the line here?
From the theoretical observations and questions discussed in this section, we conclude that finding a form of evaluation that fits this transdisciplinary research requires us to focus our approach on

the following idea: research production, the transfer of knowledge, its impact in social domains and the emergence of sustainable partnerships occur in heterogeneous networks comprising different actors pursuing distinct objectives. A successful method must do justice to what goes on in these networks, in terms of activities and performance of different actors/stakeholders.

3.4   The sci_Quest model

The above leads us to the conclusion that we are not looking for an instrument to evaluate a specific research group or program, but a process of interaction. And we are not so much looking for indicators that can tell us how good or bad the 'quality' of the research is, but we are looking for indicators that tell us whether the group succeeds in fulfilling its mission in a relevant context.

Of course we assume that a group that does not produce good quality research will not likely produce research that is relevant for specific stakeholders. Our approach, therefore, positions the unit of evaluation in the environment at large. That is, it places it in a broad array of societal domains where partnerships with stakeholders develop into more or less sustainable contexts in which agenda's are set and research strategies developed.

The method we have worked out simultaneously assesses issues of scientific and societal relevance, and therefore evaluates the diverse activities of research groups in relation to their context. The analysis weighs the different orientations of research towards its social environment. It includes feedback to the mission of a research group.

As a starting point for the method, we assume that research programs do not develop in a vacuum, but in mutual transactions with a relevant social environment. A main consequence of the interaction with the environment is that expectations, norms, values, etc. in that environment come to influence the activities of researchers and, therefore, the research output. Factors that are important in this approach are the mobility of scientists and their overall interaction with the environment (because it is essential for cross-fertilization of knowledge and know how) and the way problems are selected in such a hybrid context. We expect distinct research programs to develop in different ways and in different directions. Consequently, they might each develop a specific *profile* that is of a transdisciplinary nature. Our method sets the mission of the program up against several empirical reconstructions of its profile and its stakeholder environment.

In short, these considerations lead us to a four-step method:
1.   focus on the mission and self-image of the group;
2.   empirical construction of the research groups profile (REPP);

3. analysis of the stakeholder environment;
4. feedback phase.

To evaluate research in a reliable way, an assessment needs to be both *comprehensive* (that is, it must review all the relevant activities of the research group) and *interactive* (that is, it must allow for influence of stakeholders in the evaluation process). In the next chapter we outline the methodology we have developed to fulfill these demands. We illustrate the methodology by demonstrating several examples of its application so that the variations in its use become clear.

4 The method

In this chapter, we elaborate the four parts of our method. We not only discuss some of the different methodological fine points, but also specify the considerations behind our decisions to follow the various options we had. This might be too detailed for some readers, but we do so to show that our method implies reflection on the work and context of the research group. Our method is a framework that has to be filled in according to the specific activities of a group in relation to the relevant environment.

1   Mission and self-image of the research group

Research groups or programs differ in their mission and, we expect, accordingly in their activities and relation to the context. Since the official mission does not always coincide precisely with the group's activities and/or self-image, we therefore ask the researchers to look into the mirror and report to us what they see in terms of their performance and activities. In the first (agricultural) study we left it mainly to the groups to 'create' such a self-image by letting them attribute time-scores to different activities. In the second (pharmacy) study, we largely constructed the self-image ourselves on the basis of three different indicators that measure orientation toward several societal domains (for instance academic science/professionals, industry and policy/society) (paragraph 4.2).

2   The Research Embedment and Performance Profile (REPP)

The REPP is the empirical reconstruction of the main activities of a research group (publications, collaboration, innovation) and its performance. The REPP provides a visual representation of two critical factors:
»   the wider societal reference group for a scientific project [embedment],
»   the degree to which a project serves or does not serve the interests of the wider reference group [performance].
Data are drawn from research inputs, outputs and activities in a number of social domains, and computed into a graph. The idea is that the various activities of research groups in different social domains are depicted in a single representation (a 'profile'). The different pictures may be described in terms of different missions of research programs, for example a program can be more industry, policy or science driven. (paragraph 4.3).

3   The stakeholder analysis

After the REPP, which charts the group's internal make-up relative to the environment around it, the following step is to take a look at the research group from the outside and to consider the validation of its research by that environment: users, co-producers, client, colleagues – that is, the stakeholders. The stakeholder analysis consists of two parts, a chart of the environment of relevant stakeholders and a survey among principal stakeholders. The goal is to determine the role of stakeholders in the agenda setting process of knowledge production, in knowledge dissemination and knowledge use and applications (paragraph 4.4).

4   Comparative feedback

The fourth part of our method consists of a comparative feedback in which the results of the previous steps are analytically brought together with the research group's mission. This comparative feedback is also intended to facilitate a discussion among interested parties. Whereas the REPP shows the group's profile as a distributed orientation toward several social domains and the stakeholder analysis highlights the actual co-operation with this environment, the feedback is meant to reflect on the formulation and completion of the group's mission (paragraph 4.5).

4.1   Development of the method in practice[52]

The method we have developed contains four clearly distinguished elements. Still, it is not a standardized piece of machinery that can be applied to evaluation-work, whatever the context of the research to be evaluated. Rather, it is a flexible construct with built-in sensibility to specific contexts, allowing it to be fine-tuned to fit with the contextual contours within which the relevance of research can be evaluated. To illustrate this, we review the practical circumstances in which we developed our method. The term 'practice' is misleading in so far as it implies that 'practicing a method' is a purely instrumental and operational exercise that simply puts a methodology to work. Research and a research group, the 'object' of evaluation, are in themselves both a normative and an epistemic endeavor. As many classic studies in the sociology of science have shown (e.g. Latour and Woolgar 1979; Collins 1985; Latour 1987; Pickering 1992) every field of science is composed of both theoretical elements and socio-cultural elements (know how, experience, ways of doing, tacit knowledge). A 'practice' is the sum of these. It is not simply the empirical ways in which

---

52   In what follows, we make reference to two different studies in agricultural and pharmaceutical sciences. For the convenience of the reader we have placed one complete example from each study in the appendices. Appendix 1 shows an example of the construction of the REPP and the stakeholder analysis of a research program in the Agricultural Sciences. Appendix 2 is an example of a complete report of a research group in the Pharmaceutical Sciences.

groups go about doing things, but the locus of problematizations (Dean 1998). Following Foucault (1991) one might say that a practice entails both knowing and acting. And, while there is no exact correspondence between the two, they do interconnect and intertwine. Every evaluation process has to take this into consideration.

The acting and the knowing of researchers meet in what we call a group's 'profile'. It is not a coincidence that we prefer to speak of a profile instead of an identity of a group. The term 'identity' has the connotation of something fixed and essential. A profile, on the other hand, is the outcome of the performance of actions. It is a 'performative' (or dynamic to put it more prosaically) rather than static concept. The term profile takes into consideration that the boundaries of a scientific field (both as an organizational – by way of acting – and an epistemic – by way of knowing – domain) are open, that the actions of researchers transform those boundaries and that a comprehensive evaluation process has to take the notion of those transforming boundaries into account.

In what follows, we review how we have developed indicators and constructed our methodology in keeping with the fact that research groups transform both in terms of organizational and epistemic content and that what they do and learn has an effect on their boundaries. We then mark several steps in the evaluation process that indicate the intermingling of a group's knowing and acting. This we do that by picking up the four main elements of our approach, which indicate the different steps in the evaluation process, roughly going from 'inside out' and from 'outside in'. By 'inside out' we mean that we take the research group itself as a starting point for our analysis; looking 'outside in' entails ending with a reflection on the research group's work from the perspective of its environment (the stakeholder configuration).

4.2 Inside the research group. Mission and self image

Our approach is mission-oriented. That is, we relate the knowing and acting of a group to its self-proclaimed mission in the belief that bringing these two together can help the group focus its future efforts more successfully. The problem is that mission statements are often rather vague, sometimes only written down because some higher authority asked for them. On the other hand, missions serve as an integrative and defining concept for groups or programs. A mission is important because of its twofold effect on a group's research profile. It helps structure the group's program of action and of knowing, and it presents the group to the outside world in a certain way, thus inviting certain reactions. Creating and executing a mission, however, does not just involve looking in the mirror and saying what image you see. Like people, organizations sometimes feel it necessary to project a public image of what they take their mission to be. In other cases they might actually be rather unconscious of their mission; in some circumstances a stated mission might

not even correspond with the real face of a group (as in a kind of distorted self-image).In all cases, nonetheless, it helps structure the group's behavior and needs to be taken into account as part of the evaluation process.

Because of these complications, we decided not to rely on mission statements in the Wageningen study. Instead we asked each group to construct a self-image based on their perceived embedding in the environment, which they did by scoring their orientation to the outside world on a 100% scale. (see figures 1.4 and 1.8 in appendix 1). The result was a crystallized image of what each group took their active (performative) identity to be, one that we could subsequently hold up against and compare with reactions from outside the group. It must be said, though, that our first attempt in Wageningen was not wholly satisfactory. The evaluation committee appreciated our approach but found that the self-images were too subjective and not sufficiently comparable, making it difficult to relate them to each other. Developing a more comparable version would prove instructive.

After the first study we conducted in Wageningen, we also became more aware of the implications of taking the separate research groups (or programs) as our main unit of analysis. Instead of focusing on a faculty or university level (which is common in many other evaluations), we had decided to take the particular orientation of the research group as a starting point. Apart from the problem of subjectivity, this raised administrative questions. Since reorganizations and shifts in the organizational structure of faculties are an ongoing (and sometimes: a never ending) process in universities, we were sometimes faced with the problem that the groups we evaluated had changed faces over a period of five years, or had ceased existing. Often the head of a faculty had to be approached to resolve these administratively based problems- once again pointing to how much evaluation studies in the scientific field depend on the support of administrative management.

Apart from organizational questions, taking the research group as our unit of analysis also had theoretical and methodological implications. The main implication is that it acknowledges the variety of research programs that are gathered together in a university department. To do justice to this fact, it became clear that we needed a more objective view of each groups' specific orientation, something we couldn't get by depending solely on the self-image created by the groups themselves. We needed instead to define the groups in terms of their approach toward the scientific goals they set for themselves. This approach often appeared much broader than that found in a mission-statement, variously encompassing strategic, economic, societal and political considerations. In the first study we left it mainly to the groups to construct a self-image. In the second (pharmacy) study, we decided to construct the self-image ourselves, using three social domains in the group's environment: academic science/professionals, industry and policy/society. We still started with a group's self-proclaimed mission, but then built what we refer to as its 'global profile'; 'global' meaning not exact, a mixture of subjective and objective elements.

We did this in three (interrelated) steps. First, we asked the group to estimate what percentage of research time they devoted to actual work in the three domains, the answer to which we now refer to as the **self image** of the group. Second, we asked the group to estimate stakeholders' influence on the development of research in these several domains. This we call **contextual influence**. Third, we counted the most important stakeholders on the basis of a questionnaire and divided them over these social domains. We refer to this as the **stakeholder distribution**. Together, these three images render an idea of the group's activities in the three domains and provide a background for evaluation of its work. Formulated in a more conceptual way, these three elements represent:

1. a task distribution on the basis of external orientation of the group;
2. external influence on the research agenda;
3. the stakeholder network.

What thus appears is an articulation of the network of relationships in which the research group is situated. This network is enunciated by images: self-image, contextual image and stakeholder image. It is a reflexive rather than material concept in the sense that it is fully grounded in empirical data. In the next section we step out of the more subjective world of images and place the group in its environment in a more objective way by studying their various relationships involving texts, people, artifacts, and financing.

4.3   Inside out. The REPP

The core of our evaluation method is the construction of the so-called Research Embedment and Performance Profile, the REPP. It provides a visual representation based on quantitative data of two critical factors:

» the wider societal reference group for a scientific project [embedment]
» the degree to which a project does or does not serve the interests of the wider reference group [performance]

Where embedment refers to the position of research groups in relation to their environment, performance refers to the activities that are manifested in that environment. Together they provide an empirical profile, the visualized result of the research group's situated position and the actions that it performs.
The data that are needed to construct the REPP are drawn from several research inputs and outputs of the group. The idea is to depict a research group's various activities in a single representation. We experimented with two different graphic representations, one in the form of a radar graph (in the agricultural sciences study), another in the form of a table (the pharmacy study). Both have their own characteristics and dilemmas, as we show further on. As indicators and the differentiation

of a group's activities in several domains can be represented by various models, these variations in the construction of the REPP can be used to assess the evaluation's methodology in relation to the characteristics of the field of science being evaluated, within the conceptual framework and the consistency of the method.

Domains and indicators for embedding and performance

Any evaluation method has to consider the specific nature of a research practice. In the beginning of this chapter we defined this nature as both normative and epistemic. Both aspects are 'on the move'. Acting and knowing by researchers transform the group's boundaries. The job of the evaluation method then becomes to develop indicators that illustrate these processes. Such indicators have two functions: they have to do justice to the reality of the research that is to be described; and, taken together, they have to offer a comprehensive image of that reality. This twofold effort resembles the distinction made by Atkinson *et al.* (2002)[53] between the use and the function of 'single indicators' and the 'portfolio of indicators'. Single indicators illustrate well-defined separate developments. The portfolio refers to the whole set of indicators, the relations among them and the comprehensive image they offer. When we apply the principles for the use of indicators developed by Atkinson *et al.* to evaluation studies of research programs, we end up with the following list:

1. An indicator should identify the essence of the problem and have a clear and accepted normative interpretation.
2. It should be robust and statistically validated.
3. The indicator must be responsive to effective policy interventions, but not subject to manipulation.
4. It should be measurable in a sufficiently comparable way across research groups and comparable as far as practicable with the standards applied nationally or internationally (regarding the scale of the evaluation).
5. Indicators should be timely and susceptible to revision.
6. The measurement of an indicator should not impose too large a burden on research groups, faculties or universities, nor on the scientists themselves.

The principles applied to the whole portfolio of indicators are:

1. The portfolio of indicators should be balanced across different dimensions.
2. Tthe indicators should be mutually consistent so that the weight of a single indicator in the portfolio is proportionate.The portfolio of indicators should be transparent and as accessible as possible to scientists and stakeholders.

---

53   Although the reflection of indicators by Atkinson *et al.* was applied to the context of policymaking in the EU, the general insights are transposable to the use of indicators in evaluation studies.

With these principles in mind, we developed indicators in a number of different social domains along so-called channels of interaction. As explained in the previous chapter, we follow Callon *et al*. (1992b), who distinguish four channels: texts, people, artifacts, and financing. We consider which indicators are meaningfully applicable to each domain within each of these channels, basing our decisions partially on input from the research groups, and partially on what seems to be common in the field (which refers to some sort of benchmarking). In this approach, we refer to the communication processes and interaction patterns of a research program as the **embedding** of the program in its relevant context. The use and impact of a program are referred to as its **performance**. Embedding and performance are not completely independent of each other. Some channels of interaction represent both. The financing of contract research, for example, is both a case of a research program's embedding in certain themes of the specific context and an indication of the research's (expected) utility by those commissioning the project.

An abundant set of interaction and impact indicators and indications is available. They include: co-publications, divided research staffs, cooperation with the professional sector and the business world, contract research, professional publications, scientific articles, staff mobility, advisory positions and membership in policy platforms, involvement in special programs, publications in refereed journals and patents. Many of these indications or indicators are used in other assessment procedures for scientific programs, however mostly not in connection with each other. The most important difference between the approach used in this study and others on this point is that it does not use indicators to represent 'quality' or any of its policy-oriented variants. Rather, it calls for constructing a profile of a specific configuration of knowledge production in relation to its context, in which the program's strategic choices are apparent. A relevant set of indicators is then chosen for each of the distinguished domains, giving insight into the extent to which embedding and performance have evolved in each domain.[54] In the Wageningen study, we distinguished the following five domains:

a   Science and certified knowledge
Testing and quality control are institutionalized in the scientific community in a number of ways. One of the best known is the referee process for the acceptance of articles. There are also a number of other established, criteria-based ways to incorporate in allocating scientific reputations of excellence. Many bibliometric indicators have been developed that give insight into the spread of knowledge claims and production within the international scientific community, including a number that indicate the reception in the community in which a knowledge claim has impact. The number of publications in international scientific journals characterizes a program's scientific engagement. Membership on a scientific journal's editorial board puts scientists in the role of gatekeepers of scientific criteria. Such memberships fix a group's scientific orientation as much as the flow of

---

54   While recognizing that these indicators vary in their robustness, we do our best to bolster their indicative value.

staff to research institutions. Cooperation with other university groups is also an indication of the academic direction of research. A program's scientific reputation can be validated by financial support from a source with a more fundamental scientific signature.

b   Education and training

The production of qualified researchers is central to the domain of education and training. The number of dissertation writers and researchers in training gives a good indication of the scale of these activities and the financial support that can be gotten for taking part in them. Additionally, dissertations as a test of competence provide an indication of the yield from a course of study. Further, the external affiliation or background of students and/or researchers can be analyzed in various ways. The extent of external financing for a course of study and the extent of supplementary courses can also serve as indications.

c   Innovation and professionals

Interaction with the professional sector's domain is extremely varied. The migration of researchers to positions in business or professional (intermediary) organizations indicates a flow of skills, knowledge and influence toward this sector, as does membership in scientific advisory boards. Cooperation in projects with the business world is an indication of shared research themes. Professional publications demonstrate that knowledge claims can be translated for relevancy to the commercial sector. Financing contract research shows that commercial parties value this research orientation. Patents and royalty contracts are the most direct indications of the commercial value of knowledge and technology developed by a program.

d   Public policy and societal issues

The domain of the contribution to policy issues is the most under-developed in terms of the robustness of available indicators. There is, for example, no equivalent to the patent that is an established indicator of commercial success. Further, policy reports are not systematically distinguished as forming a category separate from other publications. Membership in organs of government and non-governmental organizations can be distinguished as well as the movement of researchers into this domain. One concrete indicator for the contribution to the policy domain is specific contract financing. It is striking that important actors in the direction of evaluative practices are rarely represented in the administrative classification used to report on that research. Although many research groups expend a good deal of their research efforts in the direction of policy themes and societal issues, the channels through which they contribute to this domain remain relatively hidden in the reporting process.

e   Collaboration and visibility

This domain was added in the Wageningen study especially to gauge the extent to which groups are embedded in the institution of which they are a part. In the case of Wageningen this was expected to be informative for the merger process that was going on at the time between the University and the DLO-institutes into Wageningen University Research Centre (Kenniscentrum Wageningen, KCW).[55] An overarching research mission guides the programs within this research center. A comparative criterion for this internal embedding can be found in the interactions with other groups in the national knowledge system and the international environment. The point here is to distinguish among these three levels of interaction in a program's environment. Indicators for this are the mobility of researchers, cooperation and the citation of publications from the program under consideration.

The Wageningen REPP: a radar graph

The radar profiles (REPP) are constructed to present the complex and multivariate activities of research groups in a way that is relatively simple, clear, complete, and amenable to comparison. The data are set out in a radar diagram as a function of a statistically average group. This allows it to show the uniformity and differences in databases. The surface indicates the domain in which a group has most strongly evolved. The surface configuration itself is the result of the mutual confirmation (corroboration, see below) of a number of indicators: the surface takes shape exclusively through the confirmation of these indicators. After a specific profile is constructed, it can be compared with other profiles.

The construction of the REPP is done with the greatest of care, but it is still a good idea to keep in mind that it is ultimately based on a particular framework with which the research began. In what follows the choice of this framework is discussed, based on a number of methodological considerations.

---

55   DLO stands for Dienst Landbouwkundig Onderzoek, a collection of applied research institutes in the agricultural sciences.

LUW programme X
% co-publ Neth (not KCW)

KCW citing
KCW co-pub
KCW coop/fin proj
% mobility to KCW
involvement NGO's (score in %)
% mobility to gov/policy arena
1,5 member gov. or spec. prog./WP
% coop/fin proj gov/spec. prog.
1,5 member advisory board/WP
2 prof art / fte total
1 patent / WP
coop/fin proj innovating inst
orientation on professionals (in %)
third badget fte (45% of fte total)

% Neth citing (not KCW)
% co-pub internat
% internat citing
% cited journ art
2 cit / journ art
2 journ art / fte total
1,5 member sci journal / fte WP
orientation on science (in %)
second fte (20% of the total)
% coop/fin proj > res. grps (incl KCW)
% mobility to research (incl KCW)
diss (cat 1) / 4 fte AIO/OIO
Junior, AIO, OIO students (1 Jun/2 Sem 1997)
a diss (cat 1) / a junior staff

Collaboration & Visibility
KCW
Public Policy
Innovation & Professional
Science & Certified Knowledge
Education & Training

% mobility to company
Research Embedment & Performance Profile

Figure 4.1 Example of REPP as a radar graph

Methodological considerations

An important criterion for choosing indicators is that they can be determined in a comparable way for every program. It is a requirement, therefore, that the measuring instruments for gathering data among groups can be applied in a stable way. Further, it should be possible to gather the necessary data in a relatively simple way. This implies that sometimes we have to rely on established categories found in the administration and reporting of research by the programs, institutes, and national science policy or external databases and bibliographical conventions. Depending on the dimensions of the assessment, it is not always possible to collect further data independently. In the case of Wageningen, it was necessary to do this through group surveys. Information about the mobility of researchers was, for example, not centrally available. Neither was information about consulting relations in the form of editorships or service on advisory boards. Additionally the survey charted cooperative links and the financing of projects. Not all the desired information was available, however. For example, memberships could only be reliably collected for the last year under evaluation.

The indicators were a method for the incorporation of the scientific and societal value of research. We selected them in such a way that they mutually confirmed (corroborated) the extent of embedding and performance. In the domain of science and certified knowledge, for example, a high average of citations was confirmed through a large output of scientific articles, relatively many editorships, considerable financial support from scientifically oriented sources, extensive cooperation with other university groups, a relatively large outpour of personnel to research

institutes, etc. Interaction channels that satisfied the criterion of uniform data were sought for each domain.

The absolute number of publications or cooperative relations does not give an adequate impression of the extent of the group's interaction with a domain. For this it is necessary to consider the group's size. In many cases, the group's research capacity (in fte's) is used for this purpose. In other cases, it is calculated according to the portion of that capacity used for relevant activities (only senior staff, for example, in the case of editorships). In other cases, an indicator is spread over the domains and the fraction of 100% is used as an indicator of embedding in each relevant domain.

In each domain, embedding and performance are constructed out of the available, collected and relevant data on interactions with that domain. Mutual comparison is one way to determine the extent of interaction or performance for a program. This is relatively simple for a number of indicators, making it easy to determine the percentage of interactions that can be allocated to a particular domain. An example is the percentage of citations by international groups in relation to Dutch groups. The 'end of the scale' is always 100%. In other cases, constructing the extent of interaction in a conceptually meaningful way is more complicated. Determining the percentage of membership on editorial boards of scientific journals in relation to the size of the scientific staff is one such example. Another method is to determine what percentage of financing comes from sources with a scientific character (operationalized here as second money stream). Both indicators operationalize the extent of embedding in the scientific domain. The conceptual problem is that it is not clear what the percentages mean. When editorships total 150% (1.5 editorship per senior staff fte) of the permanent staff's size, is it low or high? If 20% of funding comes from funds with a scientific character, how should that be related to the 150% editorship? In order to use the mutual comparison indicators, it was found that they had first to be transformed into a scale constructed with values from ZERO to HIGH drawn from an average research program. Time constraints required that this be done without systematic research; the transformation was based on informed estimates. In this case, the 150% editorship and the 20% financing from the second money stream are transformed into the score HIGH for a statistically average program in the sector under study here. (It is also possible to take a statistically average program in the Dutch research system as reference point.) The percentages 150% and 20% are numerically translated with 100 as the numerical representation of a high score for the 'average' group. All intermediate scores are similarly translated: 10% financing from second money stream funds becomes fifty on the scale of the average group and represents a modest level of scientific funding. Fifteen percent financing becomes seventy-five on the scale and represents an average level.

The numbers on the constructed scale are meaningless; only the position on the scale and the corresponding qualitative indication are conceptually significant. The translation is thus based on establishing a criterion that is valid for an average group in this sector. The fourth column of appendix 3 refers accordingly to 'end of scale'. The indicators that are directly calculated from

percentages correspond to the constructed scale without translation. The percentages are thereby labeled 25%, 50%, 75%, and 100% as low, modest, average, and high.

Research programs can be anchored in the institutional environment in a variety of ways. Some should be characterized as networks of cooperatives groups (programs to stimulate interdisciplinarity, for example), while others are thematically organized and go beyond division into departments ('afdelingen'), research units ('vakgroepen'),and chair groups ('leerstoelgroepen'), the organization of VF programs within the KCW, for example. The case of the KCW involves research programs that coincide with functional research units on a daily basis. These programs are now institutionally established as chair groups whose organizational unity was only recently introduced. The data gathering for these new institutional units is done with hindsight. All the concerned parties, however, recognize that those programs that cooperate to realize a research program (and have worked together for the past five years) are all represented in the data. This means that a program has meaningfully become a 'production unit' through its own daily management structure and has worked toward collective goals for the past five years.

In order to make a statistically reliable representation of the interaction pattern, data were assembled for a five-year period. An exception was made only for the indicators that rested on data, which were not reliable over the entire period. In such a case, the last year (1997) was taken as the standard year. The interaction profile is thus a statistical representation of the program. Given the duration of research – most projects last at least four years – it makes little sense to construct representations for shorter periods. Such reconstructions would suffer too much from indicators fluctuating due to the effects of such a short time span. Examples include generations of dissertation writers succeeding each other; the organization of conferences or having special journal issues whereby one year's output weighs heavily in the indicators. Such short-term effects need to be averaged out in order to obtain a good image of a group's profile. Additionally, it remains necessary to know something about the group's internal developments. Was there a definite break in the trend during the examined period because of the professor's departure, for example? How did personnel growth occur? Did the output strongly fluctuate; was there observable growth or shrinkage? Graphing the direct scores from a number of input and output indicators for each of the examined years can answer these questions. Such graphs make the trends easy to read.

4.4   Changes in the composition of the REPP in the Pharmacy study

An important aspect of our method is that is has to be flexible enough to fit very different kind of research fields. Agricultural science or pharmaceutical science are quite different from each other (though both can be regarded as MIT research). Therefore, with respect to the construction of the REPP considerable changes were implemented after the Wageningen study. In the pharmacy study,

the amount of publications included in the REPP per program was reduced to 15 key-publications. This implies a slightly more qualitative approach that reduces the range of the bibliometric analysis. On the other hand, this approach is more in line with the general goal of the self-evaluation, namely reflection on the group's mission. To this end, a representative group of articles may serve as an adequate sample. In the REPP we reduced the categorization of the domains in which scientific work groups are involved from five (innovation and professional, public policy, education and training, science and certified knowledge and collaboration and visibility) to three (academic (i.e. science and certified knowledge), industry (and markets) and government, policy and society. Obviously, all categories might be further refined into sub-categories such as big pharma, small industry, non pharma industry, start-ups, or as government, non-governmental organizations, patient groups and professional groups.

Communication with the scientific community about scientific developments is enclosed in the domain of 'science'. Public research and innovation process understood as the creation of innovations and competitive advantage is gathered in the domain of 'industry'. 'Participation in public or collective goods' and 'research and public debate about science and technology' are taken together in the twofold domain of 'government and society'. Next, we chose not to include education, training and embodied knowledge in one of the domains but interpreted this as a means of knowledge dissemination by exchanging people that occurs in every domain. We call this 'mobility' and analyze these movements for Ph.D. students, researchers and technicians from the pharmaceutical faculties to other scientific institutes, governmental or societal organizations, to the industry in general or to start-up companies in particular.

Furthermore, in the field of pharmaceutical research there is a particular position for the role of professionals, namely pharmacists and scientists working in academic hospitals. We categorized the former in the domain of 'government/societal' and placed the latter under the domain of 'science'. The reasoning behind this is that professional pharmacists receive courses, education and professional training by the faculties but operate in the societal domain by serving patients, while scientists working in academic hospitals (as professionals) do serve patients as well but clearly in an academic environment. Finally we had to decide how to categorize scientists working for societal organization, such as scientific referees of patient organizations who assess the research proposals of academic groups competing for funding. As became clear in our study, patient organizations clearly have two branches, one professional branch of scientists and one branch of patients. However, since the goal of this funding is to support research with a specific interest (namely the unifying 'disease' of the organization) we gathered these scientists under the domain of 'government/societal'.

Next to that, we lessened the number of indicators needed from 30 to 15 (five indicators per domain). Furthermore, we changed the graphic representation of the results of the REPP. While

in former studies we made use of a 'radar graph' we now chose to visualize the results in a less picturesque way: in a table. The table presents the relative score of the group for each indicator. It allows fewer distorted interpretations because the positioning of the indicators has no influence on the image that is generated.

The 15 key publications we analyzed in the pharmacy study were selected by the research group. We used them to review the group's impact in its relevant scientific environment. That is, the selection consisted of what the group found representative of its own output. For these publications, all citations were counted from the publication date until the date of the search. The search date is the beginning of June 2002. Apart from extracting the impact of the publications by way of a bibliometric analysis, we studied both the group's citation environment (e.g. The Netherlands, Europe, United States of America, Australia, Asia) as well as the citing institutions (e.g. industrial laboratories, governmental research institutes, patient groups).

To illustrate the impact of a shift in the visual presentation of the REPP between the two studies, we show three examples of a REPP of research groups with rather different characteristics. For each case we show a radar graph and a table, based on the same data.

Example 1. The academic picture: academic pharmaceutical science as home-ground

For each of the three social domains an equal number of benchmarks (5) are combined to construct a simple table that shows involvement and activity in each of these domains. Thus, in total 15 benchmarks represent a wide set of information on the variegated work of the group.

In the Research Embedment and performance profile (REPP)

— means a benchmark score of below approx. 50% of expected level

– means a benchmark score of between approx. 50% and 75% of expected level

= means a benchmark score of between approx. 75% and 100%

+ means a benchmark score of between 100% and 125%

++ means a benchmark score of higher than 125%

Science, certified knowledge	
relative citation impact	++
productivity: scientific publications	=/+
international visibility and collaborations	=/+
representation in editorial boards	++
invited lectures	+
Industry, market	
non-academic/commercial citing environment	+
productivity professional publications	—
involvement in industry/market	++
advisory and expert roles in commercial domain	+
editorships professional journal	no info
Policy, societal	
involvement in policy domain	=
memberships and expert roles in governmental bodies	+
memberships of societal organizations: advisory/education	+
production of public goods	=
additional grants societal/policy	++
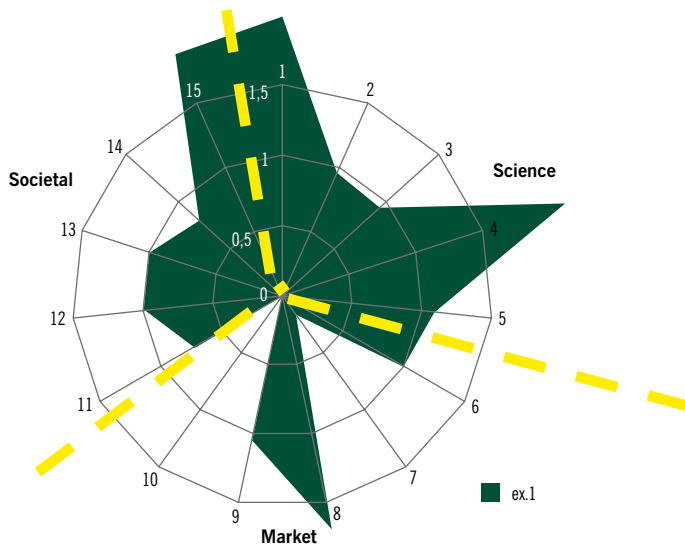
Figure 4.2 In a radar graph, the image of the group

The Repp shows a high involvement in the science domain. The scientific reputation of the group is underlined by a high representation in editorial boards, scientific prizes and serving as an invited referent for the assessment of scientific programs. The productivity of scientific publications is with 2.4 sci.publ/fte WP total nearly at the level of the reference value. The committee members should however be cautious because the reference value might be rather high for this subfield of research and the low percentage of senior staff distorts the benchmark. International visibility is good, but especially remarkable is the strong presence of a Dutch audience.

The industry and market domain is also very important, the program being involved in al sorts of commercial and 'applied' activities, and even having its 'own' firm.

The professional journals are not a very relevant category for this program and we received no reliable information regarding editorships of professional journals. The program does, however, incidentally publish in Dutch language professional journals newsletters and briefings.

The program members are quite involved in the policy and societal domain as experts and advisors. Activities are undertaken to inform and educate the general public and participation in public debate on the economic importance of pharmaceutical sciences and biotechnology. In the policy area program members are also influential in funds having the mission to promote economic spin off from scientific developments. The policy domain is thus closely related to the market domain.

Example 2. Technoscience: main orientation toward pharmaceutical science and industry; scientific excellence accompanied with considerable market relevance.

Science, certified knowledge	
relative citation impact	+ +
productivity: scientific publications	=
international visibility and collaborations	+ +
representation in editorial boards	=
invited lectures	+ +
Industry, market	
non-academic/commercial citing environment	=
productivity: professional publications	=
involvement in industry/market	+
advisory and expert roles in commercial domain	+
editorships professional journal	+
Policy, societal	
involvement in policy domain	=
memberships and expert roles in governmental bodies	+
memberships in societal organizations: advisory/education	--
production of public goods	+
additional grants societal/policy	+

The Repp shows a profile of this group with good to high involvement and visibility in the academic domain. Only the productivity with 2.37 publications/fte WP total is slightly below the benchmark of 2.5. This group's productivity should, however, be assessed by more than the number of papers published, because expert roles, patents, education and start-up firms are also an important means of producing and circulating knowledge. International visibility is good and publications are frequently published in collaboration with foreign groups. Representation on editorial boards and editorships confirm the program's position in their field. The senior members of the group are

frequently invited to international conferences as chairs. The group also organizes international conferences.

The program is also involved in several commercial activities that (more or less) generate resources to be invested in new research. Some of the contract research activities are performed by a specific contract research organization.

In the policy and societal domain this program is especially involved as an expert in governmental advisory committees. The program is also involved with societal organizations such as patient groups and grants from funds of patient organizations; it also receives additional grants.
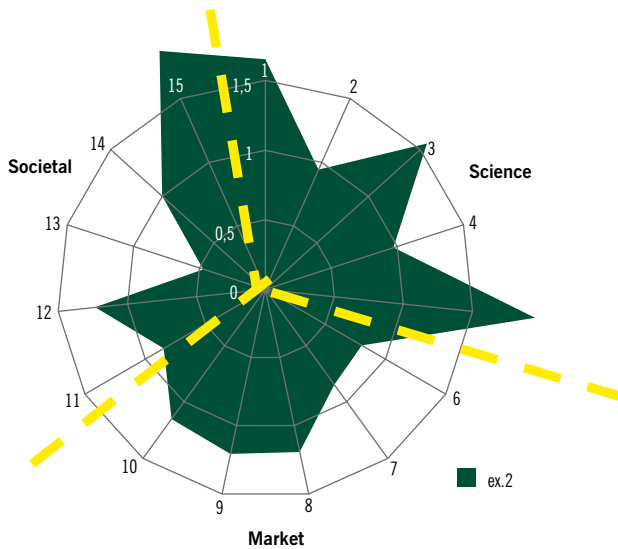
Figure 4.3 In a radar graph, the image of the group

Example 3. The societal laboratory: pharmaceutical research in a societal context

Science, certified knowledge	
relative citation impact	+ +
productivity: scientific publications	+ +
international visibility and collaborations	+ +
representation in editorial boards	=
invited lectures	+ +
Industry, market	
non-academic/commercial citing environment	+
productivity: professional publications	+ +
involvement in industry/market	+
advisory and expert roles in commercial domain	+
editorships professional journal	+ +
Policy, societal	
involvement in policy domain	+
memberships and expert roles in governmental bodies	+ +
memberships of societal organizations: advisory/ education	+ +
production of public goods	+
additional grants societal/policy	no info

This Repp shows a high profile of this program in all the domains. The Repp shows a high involvement in the science domain. Impact, productivity and international visibility are all high. The industry and market domain is not highly developed in commercial activities but more so in the involvement with a professional audience and in some projects with industry and several advisory and expert roles in this area. Patents are absent but methods for the analysis of data and databases are important 'capital' for the program. The commercial domain is nonetheless not totally absent. In the policy domain the expert and advisory roles are important, especially the memberships in advisory bodies for government and committees installed by the ministry. Projects are also

conducted with governmental bodies as partner. The program has a strong position in the social field with its knowledge on drug safety and drug policy, and also for patient groups.
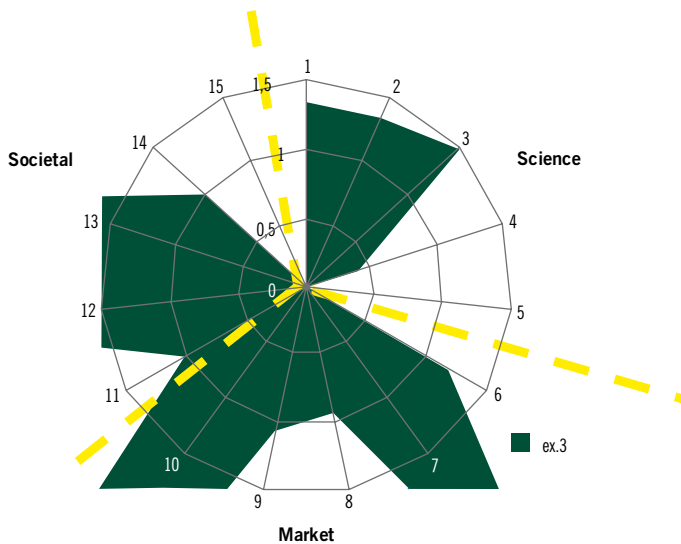
Figure 4.4 In a radar graph, the image of the group

4.5 Outside. The stakeholder analysis

The position of stakeholders

As we have seen before, an important characteristic of fourth generation evaluation is the attention given to the role of users and/or stakeholders. This attention is twofold. First, it implicates recognition of the role these stakeholders play in the actual process as it is being studied and evaluated. They are indirectly part of the innovation trajectories. Second, it means that room is created to involve stakeholders in the evaluation process.

Though the words 'user' and 'stakeholder' are generally used indiscriminately, we prefer the latter. 'User' often seems to refer to end-users only. In the case of university research, this is not very helpful. The same may even be true for research that is explicitly aimed at innovation. One tends to ignore that much innovation is the result of many different actors who mutually influence each other. A number of interested parties connected to the research can be discerned, who might not automatically be labeled users in the restricted sense of the term. While one might not speak of using knowledge products, one might yet perceive new insights arising, for example, out of a specific expertise or experience that contributes to the research or processes of socio-economic

change. The concept 'user' is therefore broadened in our analysis to include the entire field of interested parties, also known as stakeholders. Research colleagues as well as non-funding organizations with a general societal mission (such as the furthering of scientific research in a particular field, for example) are here taken into account. Research is seen here as part of an innovation process that progresses through the interaction of a multiplicity of actors, scientists, technicians, professionals, policy makers and the public. A stakeholder analysis is thus a broad inquiry that encompasses, in principle, all actors associated with innovation. This broad approach implies that the differences between actors with respect to the nature and goal of their involvement have to be accounted for.

Still, in its strict sense, the term 'stakeholder' is slightly misleading when applied to science. The term stakeholder is born in the context of companies that hold a stock-market quotation. It was introduced to clarify the question of ownership in companies. The position of shareholders in a company can be defined in terms of 'ownership' because the have capital and goods at their disposal. Stakeholders (like employees and workers) can be defined as the people who are part of the production process without owning the capital and the goods that facilitate the production.

Some of the stakeholders we identify are not part of the 'production' of science in the narrow sense, but they do have a role in the wider research process, for instance as (end)users, advisors or financiers. We use the term 'stakeholder' in a broad sense, involving all the people who have a role in the environment of the innovation process.

In another study (Spaapen 2001) we distinguished from the literature on utilization three broad groups of stakeholders, each with a different position in the research and innovation process. We used this as a basis to chart the environment of the groups. The three groups are:
1. policy makers at the intermediary or government level, whose goal is either to use research for their own policies, or to facilitate the transfer of knowledge from science to society;
2. professional users (profit and non profit); that is, industry and societal organizations that want knowledge to develop products and services (this may refer to researchers who profit from developments in other disciplines);
3. end users; that is, the public at large or individual target groups (for example farmers, aids victims).

Being in an academic environment, colleague researchers obviously form another category of users. When considering the issue of research utilization and ways to improve it, each of these groups' demands and contributions to the innovation process needs to be reviewed.

Yet, in each of these groups, there is first the question of who to include and who to leave out. The methodological principle we have used was, to paraphrase both Watergate's (in)famous informer "Deep Throat" and sociologist Bruno Latour, to follow the money, as well as the actors, articles and artifacts. Mapping the trajectories of resources that 'left' the research programs in search of their public or, conversely, the means that came in, enabled us to identify the groups' relevant environments. This was corroborated by the data from the research group who identified as part of their self-image the main stakeholders in the environment of the group. This way we constructed a stakeholder chart.

Between the two studies, we applied several refinements and changes in the categorization of stakeholders. These changes concerned classification of the kind of users, the different roles of users and the position these users have in the different stages of innovation trajectories. As we show below, in the Wageningen study we were rather ambitious in developing a model to relate the different roles of users to the different phases of the innovation trajectory in terms of 'learning environments'. The idea was to compare these different learning environments with the profiles of the research groups that resulted from the REPP. This led to some interesting results, but we were not able to make the case robust enough to continue. In the Pharmacy study, we therefore restricted the stakeholder analysis to an identification of the different kinds of users and used the survey results as a external check on the REPP.

Relation to the REPP

The analysis of the stakeholders is an indispensable part of the methodology and a necessary addition to the REPP. The REPP outlines the structure of the knowledge production in relation to its relevant societal environment. This is done through a number of selected indicators characteristic for the five different social domains that together make up the environment of the research group. The REPP is a quantification of interactions and performances. What it does not show, is whether those interactions lead to more or less stable forms of collaboration in which the interests and views of users and other interested parties (stakeholders) are represented. Thus, the REPP does not visualize the role of users in the development of knowledge production. Yet, there are important questions to be answered here. Do users have a substantial influence on the research agenda? How is the feedback from users on research and innovation organized? To address such questions, a separate analysis among stakeholders is an indispensable accessory to the REPP.

Methodological summary

The stakeholder analysis consists of two parts: a map of the environment of relevant stakeholders and a survey among principal stakeholders. The chart distinguishes stakeholders according to certain categories that can be derived from the above-mentioned distinction and adapted to specific situations. The survey focuses on the interaction mechanisms between researchers and context.

Stakeholders are asked about the kind of relationship they have with the researchers (formal and informal), about their own goals, needs and expectations, and about how they assess the particular research program. The analysis results in a description of the stakeholder environment in qualitative terms, in which we focus on the ways different actors in the innovation process interact with each other and influence the agenda setting process.

As a whole, the stakeholder analysis is thus composed of the following two parts:

Mapping
» Collection of stakeholder data (addresses, institutional background, role in the research process etc.) with the help of project lists or other information from the research groups
» Mapping the structure of the environment. Stakeholders are classified in specific domains (e.g. 'science', 'industry', 'government') according to their role and position in the research process.
» Graphic representation of the stakeholder environment

Survey
» Selection of survey respondents; usually we select a limited number of key stakeholders per research group
» Design of survey questions
» Conducting the survey (electronically)
» Interpretation of data, identification of environment
» Reporting of results

These general principals form the basis of the stakeholder analysis. In each case we altered the specific examination to make it fit the particular circumstances.

Stakeholder analysis in the Agricultural Sciences study and in the Pharmaceutical Sciences study
In the Agricultural Sciences study, the accent of the stakeholder analysis is on identifying stakeholders' roles of stakeholders and their part in the innovation trajectories in terms of learning environments. The first thing to find out therefore is what roles the stakeholders play in the innovation process. This helps in gaining insight into the course traversed from initiation of research to practical application. Stakeholders are charted in two ways, according to their institutional background and according to their financial support.

Next, the stakeholders are classified in terms of the roles they play in the innovation process. Three primary roles are distinguished: research colleagues, intermediaries and end users. This distinction in roles is needed to identify a characteristic learning environment for a particular research program. Theoretically speaking, these learning environments are characteristic for a particular phase of innovation.

The purpose of conceptualizing the learning environment in this way – and the associated characterization of innovation as a learning process – is to examine people's expectations regarding the way interaction is organized in different phases of the innovation process. For example, one might expect that a research program, which is strongly oriented toward experimental development, will cooperate more closely with a clearly identifiable group of (potential) end users than will a program that has just entered a new research area. In the second case, it is more likely that the program will have looser contacts within a broad range (government, intermediary and users) and that the contact will be directed toward advancing ideas about a desired development. Not surprisingly, stakeholders expect different things from researchers during different phases of learning. Further, a research program can function in different learning environments at the same time; findings demonstrate that this is generally the case. Differences between learning environments therefore relate to the different phases of an innovation process. Identifying a variety of possible societal and technological obstacles and possibilities is probably central in the initial phase. The learning process will be more directed toward practical application and the testing of apparatus, models, etc. as the innovation takes shape.

Following a TNO/STB study, three learning environments were distinguished in the Wageningen study (1996: 183-184). A specific form of interaction characterizes each of these learning environments. The learning environments are:

1. **Articulation:** the environment in which expectations regarding new technological developments can be expressed and articulated.
2. **Fine tuning:** the environment in which a technological nexus is created; an intermediary coordinating mechanism in which the various interests are brought in tune with each other.
3. **Test phase:** the environment in which a niche is formed and practical experiments are done outside the laboratory to test new variations.

To elaborate on the relation between stakeholders, and their part in the innovation trajectories is quite a sophisticated exercise. It is a refinement of the methodology since it relates the stakeholder analysis to a theoretical framework about the way innovation processes work. Though it offers interesting insights and a number of possibilities for further research and applications, it is not a necessary component of a stakeholder analysis. Therefore in the pharmacy study we restricted ourselves to the two main tasks (mapping and surveying).

In the case of pharmaceutical sciences, we used slightly different categories that fit that field better: (other) academic partners, industry and government/society. Obviously, all categories might be further refined into sub-categories such as big pharma, small industry, non pharma industry, start ups, or as government, non-governmental organizations, patient groups and professional groups.

Following the method, we first provided a distributed list of main stakeholders based on information from the research groups (annual reports etc.); this is the mapping. Then we conducted a survey among so-called key-stakeholders, at least two in each of the three categories. The goal was to determine the role of stakeholders in the agenda setting process of knowledge production, in knowledge dissemination and knowledge use and applications. The stakeholders were asked about their role in the collaboration, their expectations and what they see as the main goal(s) of this co-operation. Furthermore, we gathered information about the factual interaction: how they work together (exchange of people and/or equipment), what the results are (for example co-publications, joint patents, workshops, lectures, protocols etc.), and about financial arrangements.

We also asked the stakeholders to assess their influence on the direction of the scientific work, both formally and informally. We presented them a list of possibilities such as: informal contacts, participating in steering committees, advisory bodies, , conference-type meetings to harmonize interests of different stakeholders (e.g. strategic conferences, users workshops), general presentations and information services (e.g. user consultations, public debates, foresights, technology assessment research programs, opinion polls, market research), a knowledge transfer agency or courses, science shops, innovation centers, technology forums.

The survey resulted in a qualitative description of the societal position of the group understood as the relationships of the group with stakeholders from three societal domains in terms of a knowledge production environment.

4.6 Outside in. Organizing reflection on the orientation of the research group

The fourth part of our approach consists of a *comparative feedback* in which the results of the previous steps are analytically brought together with the research group's mission. The idea behind this is to focus the group on the (empirically found) profile they have developed and thereby offer them indicators and indications that can serve as means to evaluate the fulfillment of their mission. This comparative feedback is also intended to facilitate a discussion among interested parties. Whereas the mission articulates the self-image of the group, the REPP portrays the group's profile as a distributed orientation toward different social domains, and the stakeholder analysis discloses the actual co-operation with this environment, the feedback is meant to reflect on the formulation and the completing of the group's mission.

Comparative Feedback in the Agricultural Sciences study and in the Pharmaceutical Sciences study
In the case of the agricultural sciences study the focus in the feedback was on the 'learning' aspect. Similar to the work of other science researchers such as Callon *et al*. (1994) and Joly

and Mangematin (1996), a number of types of interaction profiles were determined, based on the REPPs. Within each of these categories, relations grow in a particular manner and knowledge is exchanged and develops in a specific way. Initially it was expected that a clear categorization of the groups would be possible thanks to this clustering. It was assumed that there would be three categories, similar to science-driven, policy driven and industry-driven. This division would be related to the categorical troika of the learning environment found through the user analysis: articulation, fine-tuning and the test phase. These two trios could then be compared to each other. All this was intended to give at least a heuristic indication; for example, for the phase in which the innovation process finds itself. Thus might a research group at the start of an innovative trajectory articulate its research expectations to a broad group of actors, while these actors might express their expectations in terms of technological modernization and/or implementation. At a later stage of development, political/societal fine-tuning might occur, often via intermediary organizations.

The clustering of the profiles proved not to be so straightforward. Five classes were needed to represent them adequately. It would appear that these classes were more difficult to differentiate and that they correlated less strongly with the possible phases of innovation than had been hoped. The three-part division of the learning environment also proved not to apply perfectly. In the first place, only three users were interviewed for each program, making the classification too dependent on this limited study. Further, it appeared that a number of research programs were simultaneously involved in a variety of learning environments. In these cases, the programs were classified under their dominant learning environment. This entire operation was geared toward providing the assessment commission with an easy summary of the findings that could be comparatively fed back upon the programs' missions. A good deal of information was lost in that process, leaving the resulting comparisons serviceable only as an indicative basis for the feedback process. Through a similar discussion, the future direction of the research program can be related to the mission and its existing embedment in the environment.

In the case of the pharmaceutical sciences, we put a less strong accent on the differentiation in various modes of the learning aspect. Instead, we integrated the feedback in the process of self-evaluation, a point that will be elaborated more extensively in the next chapter. The management of the faculty and research coordinators fulfilled a major role in the feedback, serving as a bridge between the research groups and us. They coordinated data and information gathering and stimulated the group-leaders to get committed to the self-evaluation. In that sense we played a more or less active role in the self-evaluation. This, of course, has pros and cons. We expect that a further improvement of the co-operation with the management and the program-leaders of the faculties may reduce the bureaucratic overhead of the operation considerably. But cooperating too closely with the people to be evaluated can also have disadvantages; for example, pressure on us to give information about other groups that would influence the process or reporting.

Learning effects

Formally, the feedback is part of the research programs' process of self-evaluation. As such it is not only the logical methodological ending of our method but also an indispensable element for the use of our method in a self-evaluation. The organization of reflection is meant, remarkably enough, to both check the robustness of our findings (does the profile indeed correspond with the self-image of the group?) and to interrogate this self-image by presenting a mirror-image. For instance, many groups have the tendency to estimate that the biggest part of their activities is oriented toward the domain of academic science, and not to industry or society. But the REPP might show very different results. This, however, does not mean that directing more attention than expected toward industry or society equals a loss of the academic scientific value of the research. In many cases these multidirectional developments go hand in hand. Most often, the interrogation of the self-image therefore is not a cruel process but a smooth correction of an image that was cherished, maintained and reproduced in academia but eroded in practice long ago. With only little exaggeration one might say that many groups proved to conduct MIT-research without even knowing it (though they *acted* in that way).

4.7 Inside out, outside in

To be short, our method focuses on action: activities of researchers and stakeholders. We try to frame the complex patterns that evolve over time in the course of action in a single graph. Around the graph, we create room for reflection (by the researchers in the self-image, and by the stakeholders through the survey).

To accomplish this, we follow the main actors. We listen to them as they articulate their mission and position themselves in their environment. We detect the streams of money, people, publications and instruments they exchange with their environment. We change from a subjective perspective in the self-image to a more objective picture in the REPP. After that, we take a reflexive turn, by asking the stakeholders to comment from their outside perspective and then go back to the group to organize a comparative feedback.

Each of the steps in the process has certain characteristics that both define its role in the conceptual framework of this evaluation method and indicate possible problems. The first part of the method (determining the self-image) and the last part (organizing feedback) are the more reflective phases. The REPP and the stakeholder analysis are the more empirical ones – the former more quantitative, the latter more qualitative. In the practical application of the method, normative, methodological and pragmatic questions have to be answered. Normative refers to the fact that many of the decisions that have to be made (such as about the amount of publications included in

the evaluation, the time period and the definition and identification of stakeholders) implicate the setting of a norm, which is never a purely scientific task. It demands deliberation with the research groups and administrators. Methodological questions arise from the fact that the approach has to be carried out in a consistent way. And because one always meets unexpected situations that have to be solved (lack of data, unwilling participants, the problem of non-response in the user survey) pragmatism is a must.

Finally, our method is comprehensive because it aims at covering all main activities of a research program or group and because it seeks to do justice to all relevant stakeholders in the research (or innovation) process.

5 Recap and future perspectives

In this final section we summarize the work we did and look to the future of research evaluation, also in relation to the newly introduced Standard Evaluation Protocol. We do so by asking a number of concrete questions that together account for our overall approach in the matter of research evaluation. These questions are:

» **What was our assignment? (paragraph 5.2)**
» **How did we interpret the assignment? (paragraph 5.3**
» **What is the gist of our solution? (paragraph 5.4)**
» **Which problems did we encounter and how did we solve them? (paragraph 5.5)**
» **What are the main conclusions and what are the options for the future? (paragraph 5.6)**

5.1   What was our assignment?

In 1998 and 2002 sci_Quest conducted studies in two fields respectively, agricultural sciences and pharmaceutical sciences. The studies were part of a longer term project of the COS referred to as 'societal quality of research' (Maatschappelijke kwaliteit van onderzoek or MKO-project). It aimed at assessing the research effort of faculties and research groups in those fields in relation to their interaction with both scientific and other relevant communities (industry, professionals, policy, the public at large). The projects were commissioned by the COS and took place under the auspices of the Association of the Dutch universities (VSNU). The faculties under study, at the Universities of Wageningen, Groningen and Utrecht, supported the projects with financial and personal contributions.

The primary goal of these projects was to help research groups and faculties find new ways of presenting their work to evaluation committees. Both researchers and administrators felt that traditional evaluations focused mainly on the so-called scientific quality of research and did not do justice to the work performed in these fields. Therefore, they were looking for methods that would represent more adequately the activities and results of their research. Sci_Quest was asked by the COS and VSNU to find a method that not only focused on the scientific quality of research, but also on the societal quality. We were to develop a method to assess research programs in their societal context. In the two projects, we experimented with ideas we had been developing in the years before and that we could now try out in practice.

The second goal of these studies had to do with feasibility. The question was whether the method we propose, including data gathering and -analysis, was straightforward enough to be performed to a large extent by the groups and/or faculties under investigation within a reasonable time and with reasonable bureaucratic effort. Implementation of this method in the national system for evaluation would only be possible under the condition that groups could make it part of their self-evaluation.

A third goal, finally, was to see whether the indicators we use in this method could be made stronger by a benchmarking operation. Though not originally intended that way, the two studies served as one long experimental track in which we altered certain characteristics, made others more visible or left them out completely. We were helped partially by the fact that the national protocol changed too.

When we conducted the first of the two studies, in Wageningen University, the old protocol was still in use. In that protocol, there was no explicit reference to self-evaluation, but it left room for experimenting with methods that evaluated aspects other than scientific quality. The new protocol, introduced in 2003, offered an excellent opportunity to investigate whether our method could be integrated into that new system. A number of the protocol's characteristics even seemed to parallel our own ideas about evaluation. The importance of self-evaluation, the accent on looking forward rather than back, the departure of comparison on the national/disciplinary level, and the discussion about the mission of the research program were all elements that – on paper at least – compare to our own approach of evaluation.

5.2   How did we interpret the assignment?

The question of societal relevance of research and how to evaluate it is getting ample attention these days, not only in the Netherlands, but also internationally. In the UK, for example, a number of studies have been conducted in the last decade to assess the relevance of research to society (Lyall and others 2004). The question has been of long time interest to the COS and, more recently, to the VSNU and the KNAW (in particular the Social Sciences Council and the Medical Sciences Council[56]) as well. These organizations have supported a number of studies in this area and remain interested in further development of theory and methods. Though the question is relevant for most scientific research, it is perhaps more obvious in some cases than in others. In the two fields we studied, agricultural sciences and pharmaceutical sciences, there are no doubts about the relevance of this topic.

The way we understood our assignment was that, in order to find a new method, we needed to develop a new perspective on evaluation, since old views generally start from the idea that research

---

56   Medical Sciences Council 2002.

is for researchers. That is, only (other) researchers can assess the quality of research, preferably the more senior people in the field.[57] This narrow interpretation does not leave room for 'societal' considerations or influence. But in order to develop a new perspective on evaluation, we first needed to develop a view on research, particularly how research operates in a societal environment.

We took as our starting point that research programs develop in mutual transactions with a relevant societal environment. The success of a research program depends on the ways in which researchers manage to connect to themes in that environment, and on the ways in which this environment absorbs ('uses') and further develops the results of the research. This starting point is supported by research in the area of science and technology studies that show that the production of knowledge in research programs is closely connected to the local organization of the research context. Researchers implicate questions and problems raised by societal actors in their programs. Such a strategy aims, among other things, at safeguarding 'resources' for research (Knorr-Cetina, 1982; Latour & Woolgar, 1979). Research programs

therefore develop in various ways depending on a particular environment. Some programs develop primarily in connection with the international scientific community (often a disciplinary community), others are more oriented toward European networks in which general policy questions are at stake, or collaborate with professional entities in a context of application. As a consequence, traditional qualifications for research groups such as 'applied', 'fundamental' or 'pure' do not fit the wide variety that can currently be detected (Crow & Bozeman, 1987).

Gibbons *et al*. (1994) underscore the change that is taking place in research and how it is perceived in their own way as they distinguish the new dynamic and organization of knowledge production from the more traditional research production with its strong orientation toward the international community. This new 'mode 2' dynamic is characterized by broad socio-economic contexts that are transdisciplinary. Mode 2 focuses in its socio-cognitive practice on accountability and reflection on what is good research. This new form of knowledge production manifests itself particularly in those areas where policy questions and commercial interests are central in the financing of research. Gibbons *et al*. notice the differences in evaluation between the two modes. In Mode 1 professional control (or certification) is predominant through the mechanisms of the scientific community. In Mode 2 other criteria are added regarding the success of research in terms of its contribution to socio-economic development and policy.

To do justice to the diversity that has developed in terms of the dynamics and organization of knowledge production, Callon *et al*. (1994) developed an evaluative instrument in which that

---

57   The classic reference here is the 1945 Vannevar Bush report that was issued shortly after the Second World War which coined the image of science as the "endless frontier". The gist of the report was that science as a whole would result in practical results, but that individual undertakings were largely unpredictable. Therefore, it was best left to the scientists to decide what to fund and what not. (see Arie Rip in Shapira and Kuhlmann 2003).

diversity is systematically represented. This so-called compass card distinguishes five domains in the context of a research program, in which separate criteria for assessment are developed. The domains are connected with particular actors and sectors. The combination of the perspective on knowledge production and the compass card helped us to develop our overall approach to evaluation that consists of four elements:

1. analysis of the mission and/or profile of the research program;
2. REPP;
3. stakeholder analysis;
4. feed back and strategic discussion.

The stakeholder analysis presents, as it were, the qualitative counterpart of the REPP. It looks at a group's interactions with its environment from 'the other side'. Since the 1970s, the literature stresses the iterative character of innovation processes (Freeman: 1974; 1991; Nelson & Winter: 1977; Nelson: 1993). That is, innovation is seen as the result of many interactions between partly overlapping social networks. But while the actions of researchers have been studied abundantly, those of the stakeholder environment have not. In the process of mutual influencing a large variety of factors such as technical, political, institutional, socio-cultural and economic (market) factors play a role. In the evolutionary literature, innovations are seen as the result of a change process in which various options ('variations') are tried out in a so-called selection environment.

The search process develops along more or less stable trajectories in which rules ('technological regimes') and social structures have evolved over time. In this development, learning processes are important because feedback from the selection environment contributes to the research and technological design. Both the learning by researchers through contacts with stakeholders and the learning of stakeholders in interaction with researchers therefore form an important element in our analysis. The stakeholder analysis aims at the different forms of interaction between stakeholders and researchers from the perspective of the former, while the REPP chooses the perspective of the research group.

The mutual exchange between the different domains is very important for the system as a whole, its viability and its form, because that exchange is its *raison d' être*. When one wishes to assess the 'innovative power' of a research program, one inevitably has to include collaborations and communications in the evaluation process. This is why our evaluation method focuses on interaction processes between research and a relevant environment.

5.3 What is the gist of our solution?

Once we were convinced that we had to include the environment in our approach, we understood that we had to focus on the relations, the interactions between research and its relevant environment. Evaluation then becomes an assessment of activities and performances rather than an assessment of quality. At first sight, that might be a disappointment to those who firmly believed that we would be able to come for the evaluation of societal quality with something comparable to what's been used for scientific quality. In fact, it is, we think, a step forward. It means that we no longer see quality ('scientific' or 'societal') as something of an absolute nature that can be expressed in a number. Quality for us is a characteristic of a relationship, it is a relative concept. (Something Diderot already expressed 250 years ago[58].)

To recap our approach we analyze the network around research programs, represented in a distribution of various social domains. Analytically, this implies a combination of two different approaches. While a domain indicates a certain well defined field or practice with its own norms, work and responsibilities, a network implies that the boundaries between domains are 'blurred'. By making use of those two perspectives we (re)construct a comprehensive image of the field of research in which we profit from the combined insights and advantages of actor-networks analyses and those anchored in domains. An actor-network analysis goes right to the variegated character of society-oriented sciences and its embedment in the context. The domain-perspective, on the other hand. acknowledges the different characters of the stakeholders involved. Although the intermingling of academic sciences with industry or other social entities can be described in terms of networks, these are very different networks with highly different partners that produce goods and demands in different ways and with a different purpose. Therefore, for the first step in the interpretation and further evaluation of society-oriented research (MIT-research), this differentiation into domains remains useful.

Still, there are interesting examples of spin-off and crossover, where the actual production of knowledge serves several domains in unexpected ways. For instance, in our study in the pharmaceutical sciences, we saw in one and the same group the dissemination of knowledge *and* tablet production technology to the detergent industry , as well as the manufacturing of an inhaler for patients with cystic fibrosis. Here, a clear boundary between the different domains of science, industry and society at a conceptual level seems to be of less importance than a method that focuses primarily on the various modes of interaction between different users and participants in the development of knowledge. Therefore, in the next step in the analysis and evaluation, a clear institutional separation of tasks is left behind.

---

58   "La connaissance est dans la perception des rapports", Denis Diderot, *Traité du beau*, Paris, 1751.

How to explain this? We think that although there are transdisciplinary practices (*networks*) to be observed, there is still a need for a clear formulation of the different tasks and responsibilities of academic, industrial and governmental and societal actors (*domains*) to deliver a vocabulary and a normative context that facilitates a further discussion about the object of academic research.

The study of Larédo and Mustar (2000) concluded that it is difficult for laboratories to be strongly involved in all activities. 'All' refers here first, to the production of certified knowledge; second, to education, training and embodied knowledge; third, to public research and innovation process; fourth, to participation to public or collective goods and fifth, to research and public debate about science and technology. This conclusion appeared to most research groups we studied and the conclusion needs to be that one should not expect them to perform well in 'all' activities. Each group develops a profile according to its mission. We have seen examples where a good performance in the domain of science is well accompanied with an equally good performance in co-operations with the industry. Also an adequate performance in the scientific domain with a good performance in the domain of government/society is no exception.

Whereas activity profiles reveal de facto strategies, they form a useful tool for evaluating the strategic choices of a lab (*ibid*. 537). Our study may be of use for that. For example, most groups show a striking lack of organized and formalized forums with their stakeholders.

Nevertheless, it is clear that research is not only directed toward the scientific domain, but develops in collaboration with other domains as well. It is precisely the extrinsic or contextual value of programs at stake in those domains that remains outside the consideration of current procedures. The radar diagrams (REPPs) that we use reconstruct the embedding and performance of research programs in the various domains that can be distinguished in their environments. They provide an image of a research program's scientific and societal orientations, which is as comprehensive as possible. The stakeholder analysis supplements the results of the REPP, especially where the expectations of the environment in terms of the research are concerned. This broadens the evaluative criteria that can be used to assess a research program and can be tailored to fit the particular program's specific mix of domains. Based on the results of both these research foci, it is further the case that the realization of a program's mission can be discussed in more detail.

5.4   Which problems did we encounter and how did we solve them?

As in every study, we encountered difficulties that were not easily solved. We list here a few of the ones we came across which need attention in further development of this method.

Choice of indicators

When developing the REPP, careful selection and consideration of indicators is necessary. It is preferable to use existing administrative and bibliographical categories to avoid excessive effort in data collection. Exceptions can be made where more detailed reporting about the research leads to more meaningful information. Along with the availability of reliable and uniform publication files, the following are of great interest: a good project administration (cooperative relations and exchange), a personnel administration (mobility of researchers), up to date records of personnel's memberships and additional jobs and a detailed financial accounting whereby an itemization of financial sources is possible. It is also desirable that the indicators be more evenly spread over the domains.

Complexity of the profile

The REPP profiles represent the various existing constellations of interactions between a program and its environment. Because of the large quantity of indicators, it is not always easy to transform them interpretively in the profile to construct such a constellation. Here one must confront the complexity of research programs and their virtually unique characters, which make it difficult to place them within a general set of classifications. This puts in danger the comparison of groups and the comparative principle of like-with-like in evaluations. It should be remarked that the act of comparison in this evaluation is, to a certain extent, artificial.

The radar surface problem

Conceptually speaking, it seems immediately clear what the surface of the radar graph means. There is, however, a complex connection between the surface and: a) the activity or performance's extent and b) the underlying confirmation of indicators (corroboration is the technical term) through the activity's completeness. The surface, in other words, is constructed out of the scores of indicators that are placed next to each other. No matter how carefully the arrangement of indicators is made in accordance with corroboration, that arrangement can prove improper for a specific group. It is also possible to arrive at a different arrangement based on other arguments. This problem can be solved by constructing the radar diagram more like a histogram (and less like a polygon). *Transformation to the construction scales* in the radar diagram appears to lead to a few differences in the sensitivity of the indicators (axes). The sensitivity of indicators must therefore be calibrated in a more systematic way. Based on research or broad consultation, it must be determined for each indicator what value is assigned to an 'average' or 'high' score for an average group. The groups can then be judged against a clear standard of reference value.

Stakeholder analysis

The stakeholder analysis is set up to supplement the REPP. As a double counterweight it aims at giving both a more qualitative insight and a view from the outside world. The biggest problem has

been the response rate. In the specific case of the Wageningen study, the number of users that could be interviewed (through telephone interviews) in the allotted time was small to begin with (30–35). This translates to only three users per research group. Especially for the larger groups, a portion of their activities remained therefore outside the picture. This was unfortunate, however, the primary goal was to offer insight into the qualitative aspects of the groups' interactions with users. In any event, we charted the users' environment as a whole and represented this graphically. In the second study, we were able to survey more people (by means of electronic questionnaires), but the answers to the questions were often not very revealing. It has been suggested that this was perhaps characteristic for the field (pharmaceutical sciences) where both corporate interests are high and intellectual property rights vital.

Critical remarks can be made about the content of the survey questions regarding the actual (end) use made of research. Some observers have said that the results of the surveys leave too many open ends. They stop where things get exciting. In the future, and perhaps this should be done via interviews, the focus should be more explicitly on questions such as what people do with the research, how the research helps users in their own practices and what precise advantage they gain. Though questions along these lines were indeed taken up in the surveys, it is clear that personal interviews can reveal more on these points. Other reasons for difficulties with the survey were that, on one side, the questions were directed too much toward aspects of interaction (and not enough toward use or utility) and, on the other side, that the expectations regarding the use of university research were not sufficiently weighted toward use in the sense of (practical) application and too much in terms of further development. This explanation does not alter the fact that in a future case more attention must be given to what actually happens with the results of research; that is, a research group's products. A combination of an electronic survey and (telephone-) interviews seems a fitting future tool for getting a more telling picture. Another option is case studies, which might be the right tool, in particular because of the time-delay between academic research and practical use. Considering the time that must be invested in this, the number of cases can only be very limited.[59]

Notwithstanding these problems, the VSNU assessment commissions recognized the supplemental value of the user analysis.

Comparative feedback
Combining the results of the two research foci in a comparative feedback is intended to evaluate the embedding and performance in relation to each program's mission, and then look forward to future research (policy) options. The challenge here is, on the one hand, to reduce complexity for the sake

---

59   Another (quantitative) option would be to look at the way the Technologiestichting STW measures the utilisation of its projects. They developed a method based on three variables: involvement of users, resulting products, and income resulting from the research. (STW 2004, p. 24-26).

of easy reference for the assessment commission and, on the other hand, not to pay too high a price in terms of loss of detail and nuance. In the Wageningen study, where we tried to classify the different programs, reviewers thought that too much in the way of programs' specific characteristics was lost in the categories that we distilled out of the constructed REPPs. But the categories did serve as an eye-opener for the singularity of the groups within the agricultural sciences sector. There also seemed to be a tension between the ideal categories (industry-driven, policy driven, etc.) that provided a sort of guiding evaluative criterion and the classification that empirically took shape from the profiles themselves. Moreover, a corresponding weighing of activities and outputs is not immediately obvious. The analytical classification used to characterize the users' environment was accepted, but the empirical foundation was considered too thin. The question was whether interviews with a few users were always sufficient to warrant generalization. Classifications require a stronger theoretical and empirical foundation. The classes can, however, serve to orient a discussion between evaluators and researchers about how to realize the program's mission empirically. For the empirical realization, however, recourse must be had to the original REPP and users' report.

In the second study we decided to be less sophisticated and leave out classifications of research groups. The focus was no longer on learning processes, but more on strategic decision making. The feedback was limited to a combination of our own interpretation of the results of our empirical investigations and the comments of the research group in question on these findings. The evaluators appreciated the information from our study as a whole. The results of the survey, that showed a clear cautiousness from the side of the pharmaceutical industry did not surprise them. The chairman of the evaluation committee, prof. J. Ruitenberg, suggested that we would have a better chance of getting more valuable information through a more qualitative approach that focused on the details of the collaboration (something we could not accomplish with our electronic questionnaire. We tend to agree with Ruitenberg that a qualitative approach, perhaps through case studies is a better way of dealing with the environment, especially in cases where sensitive interests (money, rights) are at stake.

5.5   What are the main conclusions and what are the options for the future?

The discussed assessment procedure is constructed so that the societal interests of research can be weighed in a balanced manner during the evaluation. A discussion of this methodology's feasibility cannot deal only with aspects of gathering data and method (which are treated above). It is at least as important that all the involved groups reach agreement about the initial point that the relevant contexts of research programs can differ widely and that this has consequences for the assessment procedure. While one group is more directed toward application and industry, another

group's strength is oriented toward the international scientific community. The point here is not to enter into discussions about the extent to which 'scientific caliber' can be maintained in either form of research, but only to demonstrate that differences exist. These differences are also expressed through the groups' own expectations and questions and the fact that they do not tend to reward the same forms of production. Societal relevance is central to the interactions of some groups, while fundamental scientific questions (that is, independent rather of societal utility, value, etc.) are more important to others. When evaluating a program, the needs and insights of all relevant stakeholders should be taken into consideration.

The method presented here is intended to provide the assessment commission with a systematic image of the entire context in which a group works, in relation to scientific and other stakeholders. This strongly broadens the group of involved parties who are included in the evaluation process. This approach differs in many ways from a method that uses 'objective' indicators to represent quality (bibliometrics) or a methodology that concerns itself only with the criteria of scientific colleagues (peer review). The methodology represented here strives for criteria directed toward the specific context of each research program. The point, however, is not to oppose this to other methodologies, but to broaden and supplement. It is the case that both science and government recognize that other groups in society pass judgment on the quality and relevance of scientific research, based on their own experience, and that they take part in the development of research toward application (innovation). Because both the agricultural and the pharmaceutical sciences have a long tradition of working with societal demands, it was not difficult to convince all the authorized institutions of a broader evaluation procedure's necessity and utility. More cultural resistance is to be expected in other fields.

The main conclusion is that the feasibility of our method depends on two conditions: technical and cultural (and money of course).[60] On the one hand, the technical conditions concern improving the methodological approach as spoken of above. On the other hand, the administration of research must be attuned to the demands of the data gathering process needed for this method. Culturally speaking, researchers and authorized institutions must change in the sense of broadening their assessments so that the role and judgment of interested parties other than scientists are included in the evaluation procedure.

Finally, we would like to make some remarks with regard to the goals that were set when we started the pharmaceutical sciences study, but that relate to the project as a whole. In this reflection we include the results of a meta-evaluation that was conducted by the VSNU, and the comments made by the VSNU-Review Committee Pharmaceutical Sciences.

---

60   The money remark is not completely trivial. Most researchers will say they would rather spend money on research than on evaluation (studies). However, when evaluation can be used in a self-reflective way and consequently help to improve the quality of research and the research process, it seems like money well spent. Evaluation then, instead of a periodic nuisance, should become an integral part of the research group's own policy process.

Three main goals were defined for the study:
1. to support the faculties and research groups in conducting a self-evaluation;
2. to improve the efficiency of the method: it is to be applied within a reasonable time and with reasonable bureaucratic effort;
3. to further develop the indicators and use them more strongly in a benchmarking operation.[61]

We discuss the fulfillment of these goals one by one.

Supporting self-evaluation and fitting the profiles in the system of external research adjustment
The first goal, to support self-evaluation and fit the sci-Quest method in the review procedure, has been reached by all faculties in Wageningen, Utrecht and Groningen using the methodology as part of their self-evaluations. The formal place of the sci-Quest-reports is described precisely by VSNU-Review Committee Pharmaceutical Sciences: 'The sci-Quest reports about the UIPS (Utrecht) and GRIP (Groningen) programs were presented to the review committee as "background to the self-evaluation reports". As such, they are an implementation of section A.9 of the Standard Protocol: *"In analogy with a bibliometric analysis, a methodical analysis of the institute's environment and its appreciation of the institute's conduct and results may be added".*' The Review Committee considers it to be perhaps the most important contribution of the analyses, that they "sensitise" the programs and their staff to the issue of relevance towards different stakeholders. Although the profiles were included in the self-evaluation, they were not fully integrated because the sci-Quest report was published separate from the faculties' own notes. However, we see that as a choice of the institution, fully in line with the enlarged responsibility in the renewed protocol for the review procedure.

Improving the refinement, efficiency and general application of the method
The next main goal was to refine the methodology and to simplify it in such a way that it can be applied more easily for different domains of science with minimum bureaucratic effort. This demanded changes and improvements of the methodology at several levels. First, in the construction of the REPP; second, in the stakeholder analysis; and third in how we co-operated with the faculties – that is, the organization of the interaction between the sci-Quest team, the faculty management and the program-leaders.

With respect to the construction of the REPP, considerable changes were implemented. The amount of publications included in the REPP per program was reduced to 15 key-publications. This

---

[61] These goals correspond with the questions that the COS considered to be most relevant for further exploring and refining the meaning, impact and mapping of transdisciplinary research. Its working program (2004) indicates it as follows:
1. to fit the 'sci-Quest profiles' in the system of external research adjustment as being conducted now by the KNAW, NWO and VSNU;
2. to simplify these 'sci-Quest profiles', reduce the research efforts and administrational overhead and improve the collection of data;
3. to transform the 'sci-Quest profiles' into indicators suited for benchmarking.

implies a slightly more qualitative approach that reduces the range of the bibliometric analysis. On the other hand, this approach is more in line with the general goal of the self-evaluation, namely reflection on the mission of the group. To this end, a representative group of articles may serve as an adequate sample. In the REPP we reduced the categorization of the domains in which scientific work groups are involved from five (1. innovation and professional; 2. public policy; 3. education and training; 4. science and certified knowledge; 5. collaboration and visibility) to three (1. academic; 2. industry; 3. government and society). Additionally, we lessened the number of indicators needed from thirty to fifteen (five indicators per domain). Furthermore, we changed the graphic representation of the results of the REPP. While in former studies we made use of a 'radar graph' we now choose to visualize the results in a less picturesque way: in a table. The table presents the relative score of the group for each indicator. It leaves less room for distorted interpretations because the positioning of the indicators does not influence the image that is generated.

Regarding the stakeholder analysis, it appears to be quite problematic to organize a good response from stakeholders. Especially when academic groups are tightly connected in an industrial network and have many contacts with the industry, this becomes more crucial since companies tend to be more restricted in giving information. A further improvement that can be made in this respect is to shift from a survey-format to a more qualitative approach, either through case studies or interviews with key-actors in the environment of the academic groups.

During the self-evaluation process, the management and coordinators of the faculties[62] fulfilled a major role. They functioned as a bridge between the research groups and us by coordinating the gathering of data and information and by stimulating the group-leaders to get committed to the self-evaluation. We expect that a further improvement of the co-operation with management and program-leaders of the faculties may reduce the bureaucratic overhead of the operation considerably. More learning-effects are to be expected when the data gathering and filling in the general questionnaire can be organized even more closely under the auspices of the management and coordinators of the faculties in interaction with the program-leaders.

The use of the sci_Quest profiles for benchmarking

The development of benchmarks is dependent on the availability of 'marks' that are widely accepted. This is, of course, often not the case in an area where new instruments have to be developed, such as ours. Still, we have tried to build the REPP, in the pharmaceutical case more than in the agricultural case, on the basis of benchmarks. In appendix 2 we present an overview of the determination of the benchmarks and the choices made behind them. In some cases we were able to use national figures, in others we had to rely on the insights of experts in the field. A crucial element of benchmarking is that it aims at generating a learning-effect. Instead of a mechanism

---

62   In Utrecht Dr. J. Wilting, managing director of UIPS; in Groningen Dr. H.J. Woerdenbag, scientific coordinator of GRIP and Prof. Dr. H.W. Frijlink, chair of the research program Pharmaceutical Technology, Biopharmacy and Industrial Pharmacy.

to only 'judge' the results of a group with the help of some criteria, benchmarking serves the 'coaching' of a group to help them evaluate the execution of their mission. As such, it combines the best of two worlds since it develops both objective standards ('judging') and indicators that are of help for the self-evaluation of the groups ('coaching').

5.6   Evaluating research in context: an ongoing affair

The shift from 'science to research' (Latour 1998, see also p.7 this book) that has taken place in laboratories, scientific institutions, research groups and R&D centers, and that has been proclaimed by several authors from different scientific fields (philosophy of science, sociology of knowledge, history of science) and many policymakers, policy advisors and policy watchers, is here to stay. The entanglement between societal and political problems requires a thoroughly scientific investigation, and, conversely, the societal impact of scientific innovations establishes a longstanding mutual influence and dependence. As a result of this, new questions arise concerning the validity, quality and legitimacy of a science that is so close in society. To grasp these questions, we have proposed to rename this meeting place between science and society as 'research'. More precisely, the kind of research we have helped to evaluate is MIT research, where multidisciplinary, interdisciplinary or even transdisciplinary interactions take place between science, industry, end-users, clients, co-producers of knowledge, contractors and many, many other stakeholders.

In this context, traditional procedures to evaluate research do not work because they are too one-sided. The classical repertoire of evaluation studies of science tends to regard science primarily as an academic affair, separated from the rest of the world. Without denying the strong role academic science fulfills, both in the position of an 'alma mater' and in the self-image of many researchers, there is an undeniable tendency in and around academia toward research that is strongly involved with economic, societal, political and industrial questions and that maneuvers in a strategic way. The challenge is to develop new ways to evaluate this kind of research -taking place in a very mixed setting- in a robust and transparent way.

In this book we have sketched our contribution to that challenge. We think that the four steps in our methodology are necessary building blocks for every evaluation procedure that recognizes the contextualized position of research. The exact execution and operationalization of the method may vary (though within a certain bandwidth), depending on the specific local circumstances, demands and constraints. But the following elements are crucial:
1.   a phase in which the mission of a group/program and/or its self image is established;
2.   a phase in which a more or less objective (quantitative) picture of the group's production and interaction with the environment is established;

3. a phase in which the environment is consulted about the impact of the group's work;
4. a comparative feedback, in which the results of phase 2 and 3 are confronted with phase 1, and which is meant to organize a debate on the strategy of the group.

For the first three phases we have developed indicators, benchmarks and parameters. These 'tools' have been tested in practice several times. The results of this process and the lessons that we have learned are sketched above. We will continue to study evaluation programs that fit with the current position of research in society and hope to contribute more to this subject in the future so that the evaluation of research in context may benefit from a more refined, robust and practically executable research into evaluations in context.

Literature

» Achilladelis, B. and N. Antonakis (2001), 'The dynamics of technological innovation: the case of the pharmaceutical industry', in: *Research Policy,* Vol. 30, 535-588

» Angell. M. et al. (2000), 'Is academic medicine for sale?', in: *The New England Journal of Medicine*, Vol. 342, no. 20, May 18: 1516-1518

» Angell, M. et al. (2000), 'The pharmaceutical industry, to whom is it accountable', in: *The New England Journal of Medici*ne, Vol. 342, no. 25, 22 June: 1902-1904

» Allen Consulting Group (2005), *Measuring the impact of publicly funded research,* Report to the Department of Education, Science and Training, Canberra

» Arora, A. and A. Gambardella (1994), 'The changing technology of technological change: general and abstract knowledge and the division of labour', in: *Research Policy* Vol. 23, 523-532

» Atkinson, T. et al. (2002) *Social indicators – the EU and social inclusion*, Oxford: Oxford University Press

» AWT (2006), *Ontwerp en ontwikkeling. De functie en plaats van onderzoeksactiviteiten in hogescholen*, Briefadvies no. 65, The Hague

» Bell, D. (1999) *The coming of post-industrial society*, New York: Basic Books (first issued in 1973)

» Bergmann, M. et al. (2005), *Quality Criteria of Transdisciplinary Research. A guide for the formative evaluation of research projects*, Frankfurt am Main

» Blume, S.S. and J.B. Spaapen, 'External Assessment and 'Conditional financing' of research in Dutch universities', in: *Minerva*, Vol. xxvi, no. 1, spring 1988, 1-30

» Bodenheimer, Th. (2000) 'Uneasy alliance. Clinical investigators and the pharmaceutical industry', in: *The New England Journal of Medicine*, Vol 342, no. 20, May 18: 1539-1544

» Boyer, E.L. (1990), *Scholarship reconsidered. Priorities of the Professorate*, Carnegie Foundation for the Advancement of Teaching, New York

» Callon, M., P. Larédo, T. Gonard, T. Leray and V. Rabeharisoa (1992a), 'The management and evaluation of technological programmes and the dynamics of techno-economics networks: the case of AFME', in: *Research Policy*, Vol. 21, 215-236

» Callon, M., Larédo, P., Mustar, P., Birac, A.-M. et Fourest, B., (1992b), 'Defining the Strategic Profile of Research Labs: the Research Compass Card Method', in: Raan, A. F. J. v., (ed.), *Science and Technology in a Policy Context*, Leiden: DSWO Press

» Callon, M. (1994) 'Is science a public good?', in: *Science, technology and human value*s, Vol. 19 No. 4: 395-424

» Callon, M. and V. Rabeharisoa (1998), 'Articulating Bodies: the Case of Muscular Dystrophies', in: M. Akrich and M. Berg (Eds.) *Bodies on Trial: Performance and Politics in Medicine and Biology*, Durham, N.Ca., Duke University Press

» Castells, Manuel (1996), *The Information Age: Economy, Society and Culture. Volume I: The rise of the Network Society,* Blackwell, Oxford UK

» Centraal Planbureau (2003), *Eenheid of verscheidenheid in onderzoeksagenda's? Over de bètagerichte r&d-specialisatiepatronen van wetenschap en bedrijven in Nederland,* The Hague

» Chesbrough, H. (2003), *The new imperative for creating and profiting from technology,* Harvard Business School, Boston

» Cohen, J. and C. F. Sabel (1997), 'Directly-Deliberative Polyarchy', in: *European Law Journal*, 3: 313-342

» Collins, H. (1985), *Changing order*, London and Beverley Hills, CA: Sage

» Commissie Abrahamsen (2005), *Bridging the gap between theory and practice. Possible degrees for a binary system.* Report of the Committee Review Degrees, The Hague

» Cornwall, A. and R. Jewkes (1998), 'What is participatory research?', in: *Social Science and Medicine*, Vol. 41, no. 12, 1667-1676

» Council for Medical Sciences of the Royal Netherlands Academy of Arts and Sciences (2002), *The societal impact of applied health research. Towards a quality assessment system*, Amsterdam

» *Creating an Innovative Europe* (2006), Report of the independent expert group on R&D and innovation appointed following the Hampton Court Summit and chaired by mr. Esko Aho, EUR 22005, Brussels

» Dean, M. (1998) 'Questions of method', in: Velody, E. en T. Williams (1998) *The politics of constructionism*, London: SAGE

» Deursen, P. van, F. Wamelink en Marijk van de Wende (2007), 'Kwaliteitszorg van onderzoek in het HBO: Het sci_Quest/Lect project aan de Hogeschool Utrecht', in: *TH&MA* 1 2007: 11-17

» Dijstelbloem, H. (2000) 'Overleg geslaagd, patiënt overleden – de eerste dagen van het Nederlandse aids-beleid', in: *Filosofie & Praktijk,* jaargang 21 nr.2: 3-21

» Dijstelbloem, H. (2002) 'Een plaats om bij kennis te komen. Over de maatschappelijke rol van kennisinstituten', in: *Tijdschrift voor wetenschap, technologie & samenleving,* jaargang 10, nr. 2: 34-38

» Dijstelbloem, H. en C. J. M. Schuyt (2002, red.) *De publieke dimensie van kennis*, WRR Voorstudies en achtergronden V110, Den Haag: Sdu uitgevers

» Dijstelbloem, H.O. en P. Meurs (2007) 'Leervermogen in een gemengd bestel', in: Engelen, E. en A. Hemerijck (red.) *Jaarboek Beleid & Maatschappij 2007*, Amsterdam: Boom

» Drucker, P. (1969) *The age of discontinuity. Guidelines to our changing society*, New York: Harper and Row

» Djulbegovic, B. et al. (2000) 'Uncertainty principle and industry-sponsored research', in: *Lancet* 356: 635-38

»   Edgerton, David (2004), 'The linear model' did not exist: Reflections on the history and historiography of science and research in industry in the twentieth century', in Karl Grandin and Nina Wormbs (eds), *The Science–Industry Nexus: History, Policy, Implications:* New York: Watson

»   Epstein, S. (1996), *Impure science*, Los Angeles: University of California Press

»   Etzkowitz, H et al. (2000), 'The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university-industry-government relations', in: *Research Policy* Vol. 29: 109-123

»   Etzkowitz, Henry & Loet Leydesdorff (1995), *'The Triple Helix---University-Industry-Government Relations: A Laboratory for Knowledge Based Economic Development',* in: EASST Review 14, 1995, nr. 1:14-9

»   Etzkowitz, H *et al*. (2000), 'The Dynamics of Innovation: from National Systems and "Mode 2" to a Triple Helix of University-Industry-Government Relations', in: *Research Policy,* Vol 29: 109-123

»   Fisher, D, K. Rubenson, K. Rockwell, G. Grosjean, J. Atkinson-Grosjean (2005), *Performance indicators in the humanities and the social sciences*, Centre for Policy Studies in Higher Education and Training (CHET), University of British Columbia

»   Foucault, M. (1991), 'Questions of method', in: Burchell, G., C. Gordon and P. Miller (eds.) *The Foucault effect: studies in governmentality*, Hemel Hempstead: Harvester/Wheatsheaf

»   Funtowicz, S. and J. Ravetz (1990), *Global environmental science and the emergence of second-order science*, EUR 12803EN, Ispra, Italy

»   Geuna, Aldo (1999), *The Economics of Knowledge Production. Funding and the Structure of University Research*, SPRU, Edward Elgar Cheltenham UK

»   Geuna Aldo and Ben R. Martin (2001), 'University research evaluation and funding: an international comparison', SPRU electronic working paper no. 71, University of Sussex, England

»   Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott and M. Trow (1994), *The new production of knowledge: The dynamics of science and research in contemporary societies*, London: Sage

»   Gibbons, M. and H. Nowotny (2001), 'The Potential of Transdisciplinarity', in: Thompson Klein *et al.* 2001

»   Grosjean et al. (2000), *Measuring the unmeasurable: Paradoxes of accountability and the impacts of performance indicators in liberal arts education in Canada*, University of British Columbia

»   Guba, E.G. and Y.S. Lincoln (1989), *Fourth Generation Evaluation*. Newbury Park CA: Sage Publications

»   Hagan, P. (2003), 'Review queries the usefulness of peer review', in: *The Scientist*, News from The Scientist, 4(1):20030128-05

» Habermas, J. (1985), *Die Neue Unübersichtlichkeit*, Frankfurt am Main: Suhrkamp

» Hazelkorn, Ellen, (2004), 'Growing Research- Challenges for Late developers and Newcomers', in: *Higher Education Management and Policy*, Vol 16, n 1: 119-138

» Hertog, P. den, et al. (1996), *User involvement in RTD concepts, practices and policy lessons*, Apeldoorn: TNO/STB

» IPCC report: *Climate Change 2007: Impacts, Adaptation and Vulnerability*. WG II contribution to the 4th assessment report of the Intergovernmental Panel on Climate Change of the United Nations, April 2007

» Joly, P.B. and V. Mangematin (1996), 'Profile of public laboratories, industrial partnerships and organisation of R&D: the dynamics of industrial relationships in a large research organisation", in: *Research Policy,* 25: 901-922

» Kant, A. et al. (2001) *Ongemakkelijke minnaars*, The Hague: Socialist Party

» *Kwaliteit Verplicht. Naar een nieuw stelsel van kwaliteitszorg voor het wetenschappelijk onderzoek*, (2001), KNAW, VSNU, NWO, Amsterdam

» Larédo, P. and P. Mustar (2000), 'Laboratory Activity Profiles: An Exploratory Approach', in: *Scientometrics*, Vol 47, No. 3. 515-539

» Latour, B. and S. Woolgar (1979), *Laboratory life: The social construction of scientific facts*, Beverley Hills, CA: Sage

» Latour, B. (1987), *Science in action*, Milton Keynes: Open university press

» Latour, B. (1998), *From the world of science to that of research?*, invited paper for the special symposium for the 150th anniversary of the AAAS, April 1998

» Lyall, C., A. Bruce, J. Firn, M. Firn and J. Tait (2004), 'Assessing end-use relevance of public sector research organisations', in: *Research Policy*, 33: 73-87

» Lubchenco, Jane (1997), *Entering the century of the environment. A new social contract for science*, Presidential Address at the Annual Meeting of the American Association of the Advancement of Science, 15 February 1997

» Nowotny, H., P. Scott and M. Gibbons (2001), *Re-thinking Science*, London: Polity Press

» Nowotny, H., P. Scott and M. Gibbons (2003), 'Introduction. 'Mode 2' Revisited: The New Production of Knowledge', in: *Minerva* Vol. 41:179-194

» NPRnet Conference (2002), Rethinking Science Policy. Analytical Frameworks for Evidence based policy, conference at SPRU, Sussex University, 21-23 March 2002

» Pestre, D (2003), 'Regimes of knowledge production in society: towards a more political and social reading', in: *Minerva*, Vol. 41, no. 3: 245-261

» Pickering, A. (ed.) (1992), *Science as practice and culture*, Chicago: University of Chicago Press

» Pierre, J. and B. Guy Peters (2000), *Governance, politics and the state*, New York: St. Martin's Press

» Power, M. (1997), *The Audit Society*, Oxford: Oxford University Press

» Proceedings of the International Transdisciplinarity 2000 Conference, *Transdisciplinarity: Joint Problem-Solving among Science, Technology and Society*, ed. By R. Häberli, R.W. Scholz, A. Bill and M. Welthi, Swiss Federal Institute of Technology, Zurich, Switzerland

» Raad voor Gezondheidsonderzoek RGO (2007). *De responsiviteit van universitair medische centra op vraagstukken in de volksgezondheid en gezondheidszorg.* Den Haag: RGO

» Sabel, C.F. (2004), 'Beyond Principal-Agent Governance: Experimentalist Organizations, Learning and Accountability' in: Engelen, E. en M. Sie Dhian Ho (red.) *De staat van de democratie. Democratie voorbij de staat*, Amsterdam: Amsterdam University Press

» Schmitter, P. (2001), 'What is there to legitimize in the European Union…and how might this be accomplished', paper part of the contributions to the Jean Monnet Working Paper no. 6/01

» sci_Quest/GRIP (2003), *Profiles of Research Programmes. Contribution to a comprehensive assessment of the quality of research programmes*, University of Groningen, Faculty of Mathematics and Natural Sciences

» Scriven, M. (1991), *Evaluation thesaurus*, 4th ed. Newbury Park, CA: Sage Publications

» Shapira, Ph. and S. Kuhlmann (2003, eds.), *Learning from Science and Technology Policy Evaluation. Experiences from the United States and Europe*, Edward Elgar Publishing, Cheltenham UK

» Shulha L.M. and J.B. Cousins (1997), 'Evaluation use: Theory, research, and practice since 1986', *Evaluation Practice* Vol. 18, no. 3: 195-208

» Shinn, T. (1999) Change or mutation? Reflections on the foundations of contemporary science, in: *Social Science Information,* Vol. 38, no.1:149-176

» Snow, C.P. (1959), *The two cultures and the scientific revolution*, Cambridge University Press, New York

» Spaapen, J.B. (2001), 'Utilization of research in North and South. A review of recent literature', in: *Utilization of research for development cooperation. Linking knowledge production to development policy and practice*, RAWOO publication no. 21 (Advisory council for development policy), The Hague

» Spaapen, Jack and Christian Sylvain (1993), 'Assessing the value of research for society', in: *Research Evaluation*, Vol.3, no.2, August 1993: 117-126

» Spaapen, J.B. (1995), *The Evaluation of Research for Society*, dissertation, University of Amsterdam

» Spaapen, J. and F. Wamelink (1999), *The evaluation of university research*, NRLO report 99/12[E]

» Shinn, Terry, Jack Spaapen and Venni Krishna (1997, eds.), *Science and technology in a developing world*, Kluwer Academic, Dordrecht

» Shinn, Terry (2002), *'The Triple Helix and New Production of Knowledge: Prepackaged Thinking on Science and Technology'*, *Social Studies of Science*, Vol. 32, No. 4: 599-614

» Social Sciences Council and Humanities Council (2005), *Judging Research on its Merits*, Royal Netherlands Academy of Arts and Sciences, Amsterdam

» *Standard Evaluation Protocol 2003-2009 for public research in the Netherlands*, KNAW, NWO, VSNU (web publication only)

» Stokols, D. et al. (2003), 'Evaluating Transdisciplinary Science', *Nicotine and Tobacco Research*, Vol. 5, supplement 1, s21-s39

» Stokols, D. (2006), 'Towards a Science of Transdisciplinary Action Research', *American Journal of Community Psychology,* 38: 63-77

» SWR (1983) *Beklemmend Wetenschapsbeleid*, werkdocument nr. 8, Noord-Hollandsche Uitgeversmaatschappij, Amsterdam

» Technologiestichting STW, *Utilisatierapport 2004*, NWO, Den Haag

» Thompson Klein, J. (1990), *Interdisciplinarity. History, Theory and Practice,* Wayne State University Press, Detroit

» Thompson Klein, J., W. Grossenbacher-Mansuy, R. Häberli, A. Bill, R.W. Scholz, M. Welti (2001, eds.), *Transdisciplinarity: Joint Problem Solving among Science, Technology and Society. An effective way for managing complexity*, Birkhäuser Verlag, Basel

» Thompson Klein, J. (2003), 'Bridging Research and Social Interest: The challenges of evaluation in transdisciplinary research and public policy', Unesco-Most Regional School for Latin-America and the Carribean: Local development and governance, Punta del Este, Uruguay, 10.28.2003

» United Nations, *The Secretary-General's report 2000 for the Millennium development Goals*, New York, 2000

» Vannevar Bush, *Science The Endless Frontier A Report to the President on a Program for Postwar Scientific Research*, July 1945

» Verkaik, A.P. (1997), *Uitdagingen en concepten voor toekomstig landbouwkennisbeleid*, NRLO report 97/17, The Hague.

» Vries, G. de (2001), 'Medische wetenschap is steeds afhankelijker van commercie', in: *de Volkskrant* 22 februari 2001

» Weatherall, D. (2000), 'Academia and industry, increasingly uneasy bedfellows', in: *Lancet* 355: 1574

» Werkgroep Kwaliteitszorg (2001), *Kwaliteit Verplicht. Naar een nieuw stelsel van kwaliteitszorg voor het wetenschappelijk onderzoek*, KNAW, VSNU, NWO, Koninklijke Nederlandse Akademie van Wetenschappen, Amsterdam

» WRR (2004) *Bewijzen van goede dienstverlening*, Rapporten aan de Regering Nr. 70, Amsterdam: Amsterdam University Press

» Wilts, A. (2000), 'Forms of research organisation and their responsiveness to external goal setting,' in: *Research Policy*, Vol. 29, no. 6 June: 767-781

Appendix 1
Case example agricultural sciences

Introduction

Crop and Grassland Science, with an average input of approximately 5,5 fte 'WP total', is one of the larger programs in our sample and more or less similar in size to its 'sister program' in Agronomy: Plant Production Systems. The program is the result of a merger (in the reorganization process in 1995) of two initial 'chair' programs: 'crops' (Struik) and 'grassland' ('t Mannetje). The Input figure presented here shows that a result of this process was a slight decline in total input, though not in tenured staff (WP1). A sharp decline is found in the number of Ph.D. students - AIO+OIO's (we will come back to this in the discussion of 'Education and Training'). In positioning the research program, emphasis has been laid on the provision of scientific information for quantitative understanding of crop productivity. Disciplinary expertise is stressed and steps are taken to enhance this disciplinary expertise.

1   Output

Publication output of the program does not systematically show the effects of the decline in input. Output fluctuates slightly between categories but remains generally stable with the exception of 1994. At that time, the 17th Int. Grassland Congress in Palmerston, New Zealand and the 15th General Meeting of the European Grassland Federation in Wageningen boosted the production in conference proceedings. In 1995 and 1996 articles found in the SCI/SSCI exceed the reported articles in refereed journals. CGS is co-publisher of these articles, but they were eventually uniquely assigned to another program in the annual report of the university. Overall, we conclude that the portion of journal publications in SCI/SSCI is substantial and growing.

2   REPP

General impression of the profile is that activity and performance in the domain of Science and certified knowledge is comprehensive at an average to high level. Exception is funding in the second money flow. Education and training is exceeding 'end-of-scale' on the two indicators. The criterion weight of 1 dissertation per 4 fte. Ph. D. Student is exceeded by a huge margin. Involvement in Innovation and professions is only partially significant. Public policy is relatively low. In the domain

of collaboration and visibility substantial interactions with other programs of the KCW on all indicators spring to the eye and even a higher lever of citations from foreign programs. Interactions with other Dutch programs are almost absent.

KCW

CGS structurally collaborates with the DLO Centre for Agro-biological Research and frequently with the department for Theoretical Production Ecology; the department for Farm Management and the department for Plant Physiology. This reflects KCW policy to stimulate cooperation between research programs. Furthermore CGS collaborates with the Research Station Arable Farming and Field Production of Vegetables in Lelystad. International collaborations are lower as compared to the KCW, and more scattered. The International Centre of Agricultural research in dry areas in Syria, the Scottish crop research institute in Dundee, Scotland and the department Agronomy of the University for Agricultural Science in Bangalore, India are the only programs being more than an incidental partner in SCI/SSCI publications. In the citing environment, the number of citations by foreign programs is higher than those made by programs from the KCW. International visibility is broad both geographically and in orientation of the citing programs but also mainly incidental.
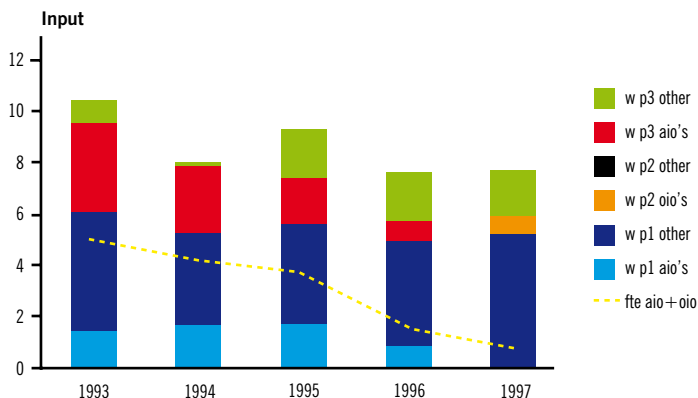
Figure 1. REPP Crop and Grassland, appendix 1.1

Output

professional publications and report
other scientific publications
conference proceedings
chapters in books
books
dissertations cat. 1
journal articles
– – – # criteria journ. art. (period till '98)
– – – sci/ssci articles
– – – citations journ. art. in next year
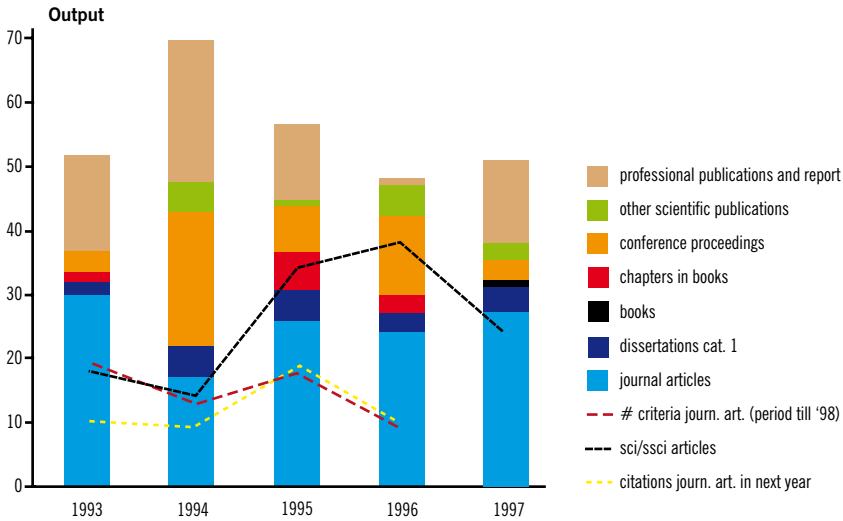
Figure 2. REPP Crop and Grassland, appendix 1.2

Crop and Grassland Science
% co-publ. Neth. (not KCW)

KCW citing
KCW co-pub.
KCW coop./fin. proj.
% mobility to KCW
involvement NGO's (score in %)
% mobility to gov./policy arena
1,5 member gov. or spec. prog./WP
% coop./fin. proj. gov./spec. prog.
1,5 member advisory board/WP
2 prof. art./fte total
1 patent/WP
coop./fin. proj. innovating inst.
orientation on professionals (in %)
third badget fte (45% of fte total)

% Neth. citing (not KCW)
% co-pub. internat.
% internat. citing
% cited journ. art.
2 cit./journ. art.
2 journ. art./fte total
1,5 member sci journal/fte WP
orientation on science (in %)
second fte (20% of the total)
% coop/fin. proj. > res. grps. (incl. KCW)
% mobility to research (incl. KCW)
diss. (cat. 1)/4 fte AIO/OIO
Junior, AIO, OIO students (1 Jun/2 Sem 1997)
a diss. (cat. 1)/a junior staff

KCW
Collaboration & Visibility
Science & Certified Knowledge
Public Policy
Innovation & Professional
Education & Training

% mobility to company
Research Embedment & Performance Profile

Figure 3. REPP Crop and Grassland, appendix 1.3

Science and certified knowledge

One of the parameters for embedment in the broader scientific environment is the citation. Together, citations present an image of the interest of other researchers around the world in the work of a program, or of the 'visibility' of the program. The average publication of CGS receives 0.49 citation per year. To estimate scientific merit one has to keep in mind that there are major differences at this point between fields. We therefore supply the reader with some rough perimeters that may serve as an indicative yardstick. The average article in the best cited agricultural journal receives approximately 1.7 citations. In Potato Research, being an important outlet for this program, articles receive on average approximately 0.2 citations yearly.
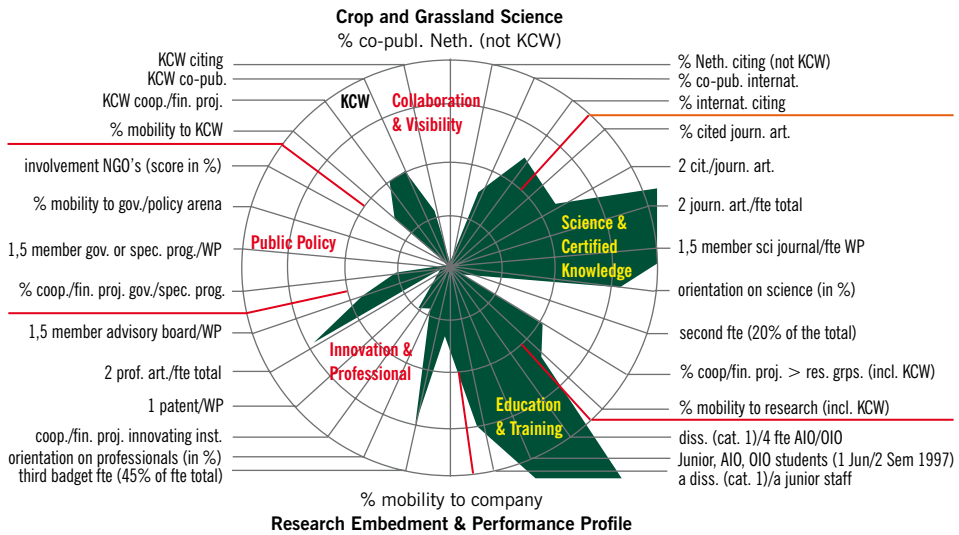
The Netherlands Journal for Agricultural Sciences receives approximately just below 0.5 citations on average and thus would compare almost precisely with the overall visibility of the CGS publications. In general, publications are made in agricultural journals, a few are made in more traditional disciplinary journals. Production of articles in refereed journals comes close to 3 journal articles / fte total (average = 2.86), making CGS the second most productive program in the publications of journal articles. Members of the program are well-established in editorial boards of scientific journals and series. Respondents indicate an orientation on science as high as 80%. Funding in the second money flow is only recently acquired (see figure 'input').

Substantial interactions with scientific actors are confirmed by the over 50% projects in which the program cooperates with public research programs (most of them KCW programs) and the mobility of over 50% of personnel leaving the program for another job in research.

Education & training

With a production of 19 dissertations, while only 15 fte AIO+OIO are reported (expected 3.75 dissertation), this program scores high on production of dissertations in other Ph.D. arrangements: e.g.: 'sandwich Ph.D.'s (extra 15.25 dissertations cat. 1). This is not completely unusual in the Agricultural University. Part of this high production may be explained by higher investments in Ph.D. students in previous years. We found a sharp decline in the portion of AIO+OIO (Ph.D.) students (see figure 'input'). Second explanation may be the 'sandwich' Ph.D.'s construction. On the one hand, the effort to supervise these Ph.D. students must be considerable. On the other hand, also Ph.D. students contribute to the production of the program. A good illustration is the response of this program to the question for the most important products having societal and technological value: five dissertations and related publications are mentioned. On average, the ratio fte Junior program member (1997) to fte Senior program member (1997) is just over 0,5 fte Junior to 1 fte Senior. This is what we considered to be a 'modest' involvement in training of juniors. This does not represent the involvement in the training of juniors over the whole period (see sharp decline in AIO+OIO). For this total period, one has to conclude involvement in education and

training has been rather high. The number of dissertations compared to the number of junior staff members started before 1996 suggests some delayed unfinished Ph.D. projects.

Innovation and professional domain
CGS is clearly present in the innovative and professional domain. Approximately one third of funding originates from contract research; professional publications are one third of the total of scientific and professional publications aggregated.

Average journal article per fte WP	Average journal article / fte WP
Vegatative propagation	1,61
Technology and Agrarain Development	1,24
Soil Tillage	0,69
Systems & Control	2,02
Plant Production Systems	1,60
Irrigation and Water	
Engineering Program	0,42
Greenhouse Horticulture	1,59
Farm Technology	1,35
Erosion and Soil & Water Conservation	0,92
Crop and Grassland Science	2,86
Applied Physics	2,97

Most important contractors, however, are the Dutch ministry for Agriculture, Nature Conservation and Fishery (LNV) and the European Union (EU), who both belong to the policy domain. This is reflected in the cooperation and funding of projects which is approximately for twenty five percent with innovative and professional organizations and slightly less with governments and special programs. In projects CGS is structurally interacting with the research station Arable farming and field production of vegetables in Lelystad and is partly setting the research agenda of this station. Furthermore several contacts exist with commercial firms e.g. in the context of the Aloë project in Aruba and CGS contributes to seed-improvement. Program members occupy advisory functions in the sector of potato-firms, industry interested in application of hemp and the Nutrient Management Institute. Extension of the hemp project in an EU network project, for example, suggests interest in

this competence by other actors. Also new production methods are developed in activities defined as Research and Development.

Public Policy

Project funding and participation is only slightly lower compared to the innovative domain. The program however is stressing its relevance in this domain in the light of EU policy on reducing nitrate levels in soil water and its relevance for the development of national legislation on nutrient management on farms. Existing competence on the modeling of crops seems to find an area of application in the domain of public policy. Contracts resulted from this and reports have been presented to the ministry of Agriculture (LNV). Embedding however seems to be only partially developed at this moment. In addition to that, the program is defining itself mainly as scientifically oriented. In defining the 'embedment in the environment' the program is strongly emphasizing interactions with the KCW, followed by Dutch government, Dutch funding and international non-agricultural research institutes and funding. Commercial actors are missing and only Dutch central government is present in the policy domain. From this, the conclusion can be drawn that the program is predominantly evolving in the scientific and the educational regimes. Involvement in the sector of innovation is clearly secondary.
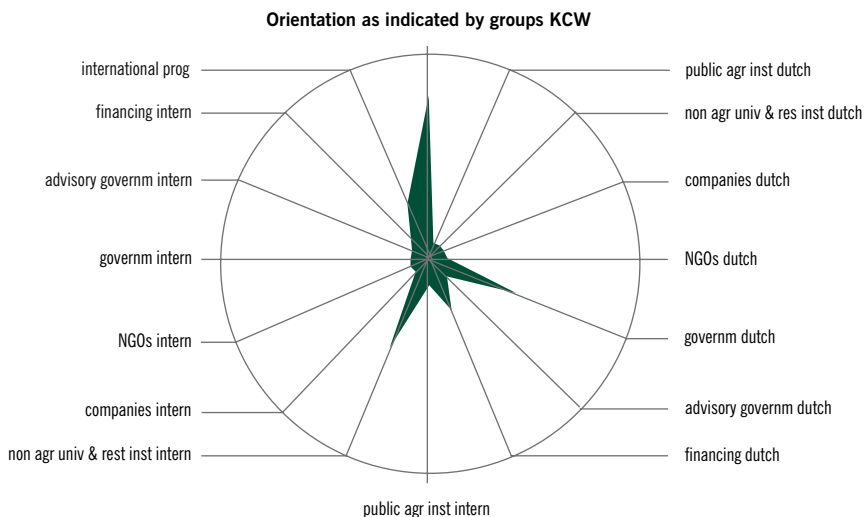
Orientation as indicated by groups KCW

Figure 4. Appendix 1.5

3　The stakeholder analysis

Respondents
Ministry of Agriculture; Mommersteeg international; ADAS consulting UK.

Research mission
The objective of research is to provide information which is necessary for a quantitative understanding of crop productivity, to provide the scientific mass for the development of better tools for crop and grassland management and to integrate the knowledge in designs of plant ideotypes and designs of improved, more sustainable cropping systems.

Role/function
Two of the three users here relate to the group through relatively small projects (14,000 – 27,000 euros). ADAS, a British consultancy firm is coordinator of a much larger European project, 'Hemp for Europe'. They see their role as both coordinator and researcher. The Dutch user from industry sees his role as commercial developer (obviously), but also as colleague for the research group. The policy maker perceives his role in an uncomplicated manner: there was a specific question and the right group to answer this question was needed. Both the national policy maker and the industry respondent had knowledge of the research program through personal contacts (which in itself is an indication of knowledge transfer).

Interaction mechanisms/arrangements
Contacts are arranged in rather different ways between the group and these users. With the national policy maker, contacts were incidental, mainly due to the small size of the project.
With the Dutch industry, contacts are much more frequent, also people are exchanged on a regular basis (for other projects as well). Clearly, this has to do with the fact that joint experiments are conducted. Interaction in the European project is the most extensive of all. There are steering committees in which users participate, specific meetings where involved actors attune their interests, and experimental test sites. The goal of the interaction is for all three users to find answers to a specific question (which makes the policy maker somewhat atypical, since usually demand is oriented towards more general policy issues). For the British consultancy firm, the goal of the interaction with this group is much more: developing new technology, input of specific expertise in the research process, and diffusion of new technology. Clearly, this is largely due to its role as coordinator in a big European project.

Specific research methods
The industry participates in the research process through the providing and harvesting of seeds. The national policy maker keeps at a distance, not so much because it is a general policy rule, but

because it is such a small project. The British consultancy firm runs a complimentary experiment, and coordinates the project as a whole. Neither of these users has had experience with specific methods to involve users (that is, apart from jointly conducting experiments).The group proper mentions 'input of growth technical data in economic modeling' as a user oriented research method.

Types of use

The output that is used by each of the users is rather specific. The policy maker needed numbers for a specific part of a government policy (manure regulation), and the group was able to provide that. The industry was interested in method development for their own commercial activities.
The consultancy firm wanted knowledge for specific experiments. Two of these users do not feel extremely dependent of this group, since others have similar knowledge (CPRO, IRS).
The third one (ADAS) calls the input of the group 'essential to the effective running of the project'. Potential users of the work of this group are the ministry of agriculture, and farmers.

Overall assessment

The overall assessment by these users of the interaction with the group is good. Two users praise the reliability of the group and the quality of the work. The third one (ADAS) the complimentary technical skills. The connection between research and demand was good, the possibilities for users to influence the research agenda sufficient, and there is no hesitation with regard to future cooperation. The coordinator of the European project refers to this group as 'one of the best partners in the project.'

Learning environment

Clearly, this group is in category C. with industry, and within the European project, experimental work is conducted. The policy maker was provided with prompt answers to a specific question.

Appendix 2
Case example pharmaceutical sciences

Introduction

Research groups in the pharmaceutical sciences operate at the crossroads of science, industry and society and policy. This means that activities of researchers are not only geared towards scientific colleagues, but also to other stakeholders. In the following paragraphs, this broad array of activities is presented through a number of quantitative and qualitative indicators. Data were gathered in three different ways: a questionnaire filled in by the research group, an email survey among stakeholders, and bibliometric data. Additional information from the annual reports and from the faculty (input figures) was also used. The group will be presented in the following way: First we will start with a description of the self-proclaimed **mission** of the group, followed by a **global profile** of the group (i.e. the way the group relates to its environment). In section three the group is presented in terms of a number of **input and output** indicators; section 4 is a limited **bibliometric analysis** based on the group's 15 most relevant articles over a five year period. In section five a comprehensive picture is given of the group's activities in three social domains: science, industry, and society/policy. This is referred to as **REPP**: research embedment and performance profile. In section six, an analysis of the **stakeholder environment** is performed. Finally, in section 7, the results of the previous sections are summarized and related back to the mission of the group.

1   Mission of the group

The Social Pharmacy and Pharmaco epidemiology group has its focus on the social dimension (in a broad sense) of the utilization of drugs. Therefore, it is closely involved with professional organizations (pharmacists), societal organizations and policymakers. According to its own mission statement, the group conducts basic research in order to assess benefit-risk profiles of drugs, and to understand the mechanism of the optimal application of these drug profiles in pharmaceutical care. Major areas of research include pharmaco epidemiology, drug utilization, pharmaco economy, drug information and patient education as well as business and science studies. The aims of the group are to generate and optimize dynamic databases and computational tools for the assessment of benefit-risk profiles of drugs, to implement these benefit-risk profiles of drugs in protocols for optimal pharmaceutical care in the health care setting, and to evaluate the effectiveness of these protocols in pharmaceutical care.

2   Global profile of the group

We would like to present a global profile of the group in terms of its relations to the three main social domains: academia, industry and policy/society. We accomplished this by using three separate questions. First, we asked the group to estimate what percentage of research time is devoted to actual work in the three domains (**self image**). Second, we asked the group to estimate the influence of stakeholders in the three domains on the development of research (**contextual influence**). Third, we counted the most important stakeholders mentioned the questionnaire and divided them in the three social domains (**stakeholder distribution**). Together, these three images render an idea of the group's activities in the three domains and thus a background for evaluation of its work.

Self image

This group's research orientation seems to reflect its mission accurately in connecting fundamental research to the study of practical drug utilization and its societal implications. Orientation towards industry is relatively small, whereas the contacts with professional organizations (pharmacists, health care) are many. Estimates are in percentage of total research time.
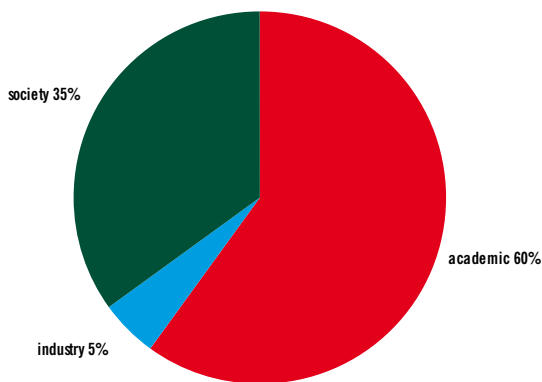
Figure 5. Research orientation in three social domains

Contextual influence

When we compare the self-image with the way the group estimates the importance of partners in its societal context, we see a similar picture, but more extreme. Industry accounts for only 5% external influence, the government/society sector for 20% and the academic sector for 75%. It must be noted that here the government/society sector refers exclusively to government. So, though the group collaborates with several societal partners, the influence of such partners on the

research program is apparently non-existent. It would appear that 'collaboration' here refers largely to applied research, that is applying knowledge and/or expertise in certain practices, for example in monitoring the use of a particular drug. This is in any case main goal of the group; specifically that is the implementation of research results into the pharmaceutical practice. [At the same time, the group states that in these applied studies there is always interaction with a particular intermediary, often a representational actor of the pharmaceutical or medical practice]

Stakeholder distribution

When we look at the main stakeholders in the group's environment, we see a rather different picture. Here, the society/government sector scores the most: 12 collaborative actors compared to 8 in the academic sector and four in industry.

society 50%

academic 33%

industry 17%
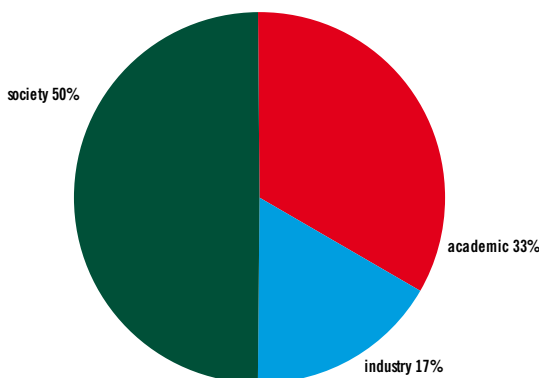
Figure 6. Main stakeholders listed by social domain

3   Input, output and mobility

In the following graph the relation between input (money and personnel) and output (products) is given between 1996 and 2001. Note that the definitions of publications differ from other Groningen programs, as this program distinguishes between international and national scientific publications. The labels in the graph correspond with this distinction.

The program is growing in all respects. Input doubled and the output also follows overall in this development. The program does publish a high number of scientific publications, nearly the double of the benchmark of 2,5 sci publ/WP fte. In this case we have to note the very high percentage of senior staff (more than double that of the average program). Correction of that would bring the

productivity at a slightly below average level. Also the production of professional publications, mainly in the Dutch language, is high. These include broadly appreciated professional journals such as the *Pharmaceutisch weekblad, Nederlands Tijdschrift voor Verloskunde, Ziekenhuisfarmacie, Infectieziekten Bulletin and Apothekersvademecum*. This suggests that the program is very active in the broader diffusion of knowledge to professional groups. The 'other publications and chapters of books' can also be defined as professional publications. The number of dissertations is growing. Contract research, although fluctuating over the years, can still be defined as a stable trend. The program reports that some of the changes visible in the graph before 1999 might partly be due to a less consistent classification of publications.
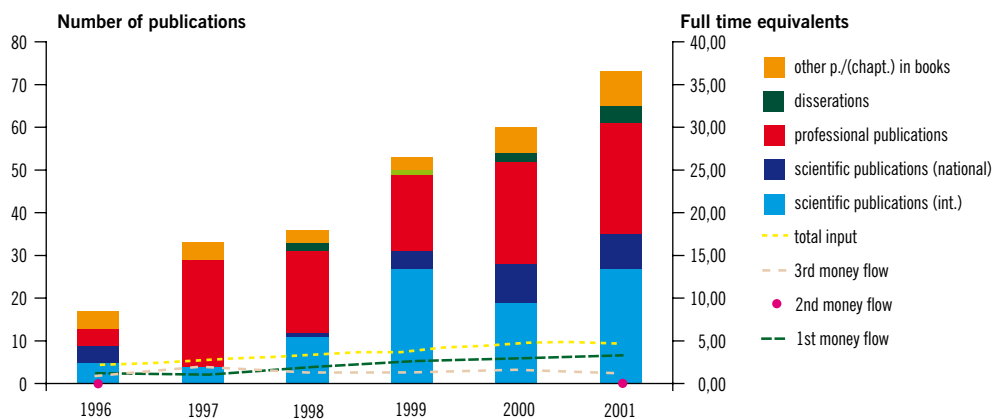

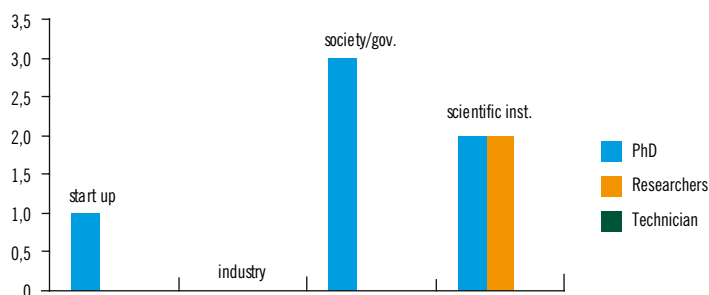**Figure 7.** Input-output 1996–2001

Figure 8. Mobility during 1996–2001

Mobility during the period 1996–2001 (note that the group officially started in 1998) was mainly in the category of PhD students. Of the six that moved away, three went to the public sector, two took a job as a researcher in other scientific institutions, and 1 started his/her own company. In that period two staff members went to another scientific institution.

4   Impact of the group in terms of 15 key publications

To review the impact of the group in its relevant scientific environment, we analyzed 15 key publications, selected by the group. That is, the selection consists of what the group finds representative of its own output. For these publications, all citations were counted from the publication date until the date of the search. The search date is the beginning of June 2002. In table 1 the most important straight data are presented. This program, being young and growing, selected nearly half of the publications from 2000.

These publications are rather recent in attempting to establish their impact. For example, five publications have not (yet) received citations. The number of citations is also highly influenced from self-citations by the authors. This indicates a rather coherent group of researchers referring to their own work. The impact of the publications remains behind the average number of citations in this journal set and the average number of citations of the citing set. This might have to do with the specificities of this particular field. Publications appear in medical journals with a distinguished reputation, but because they are epidemiological and social pharmaceutical publications this results in the fact that they do not receive on average a comparable high number of citations. The publications can be defined as 12 articles, one review article (Drug Aging), one letter (Lancet) and one editorial (Drug Safety). The later two publications, as a rule, receive fewer citations than the other types of publications.

Table 1. 15 Key publications

Journal	publication year	volume	page	citations	all authors self citations	in-group citations	impact factor 2000	half-life 2000	# citations 2 yrs window
Contraception	1998	V57	247-249	1	1	1	1,704	7,100	1
Brit J Clin Pharmacol	1998	V46	255-261	2	1	1	2,151	7,800	2
Drug Safety	1999	V21	153-160	0	0	0	2,763	4,500	0
Brit Med J	1999	V319	291-292	4	1	0	5,331	0,000	4
Lancet	1999	V353	1187-1187	6	2	0	10,232	6,900	2
Brit J Clin Pharmacol	1999	V48	239-246	2	2	2	2,151	7,800	2
Eur J Clin Pharmacol	1999	V55	139-144	12	6	1	1,729	8,800	10
Teratology	1999	V60	33-36	2	1	1	1,600	10,000	1
Brit J Clin PharmacoL	2000	V49	254-263	2	1	1	2,151	7,800	2
AIDS	2000	V14	2383-2389	0	0	0	8,018	3,400	0
Drug Aging	2000	V17	217-227	0	0	0	2,342	3,900	0
Eur J Epidemiol	2000	V16	329-336	0	0	0	0,918	5,900	0
Brit J Clin Pharmacol	2000	V50	473-478	1	1	1	2,151	7,800	1
Drug Safety	2000	V22	321-333	6	3	3	2,763	4,500	6
Paediatr Perinat Epidemiol	2000	V14	111-117	0	0	0	0,000	0,000	0
					50%	29%			

Table 2. Citing environment

Citing address	Citing %
IN-GROUP	24%
Netherlands	41%
Europe	54%
North America	11%
Asia	3%
Australia	3%

The citing environment of this program is dominantly Dutch and European. The high number of Dutch institutes in this field indicates a strong network within the Netherlands. Collaborations within the selected publications also indicate well-developed relations with several Dutch hospitals, municipal health care organizations (GGD's), governmental laboratories (RIVM, TNO) and a few international academic groups. The international visibility is modest, compared to other programs, however, collaborations with international groups are relatively good.

Table 3. Citing institutes (rank ordered by citing frequency)

Citing Institute	Country
University Groningen	Netherlands
Aalborg Hospital	Denmark
Aarhus University	Denmark
University of Texas	USA
University Nijmegen **(ST Radboud??)**	Netherlands
Aarhus University Hospital	Denmark
Ctr Dis Control & Prevent, Birth Defects & Genet Dis Branch	USA
University Nijmegen Hospital	Netherlands
Univ of Oslo	Norway
Tweesteden Hospital	Netherlands
General Hospital De Tjongerschans	Netherlands
Chu Angers Hospital	France

The table records the first part of a long list of institutes citing the group's publications. The institutes are rank ordered by frequency of citations. It breaks off at a citation frequency of three citing articles (usually the list is presented until five citations). Industry groups are not present in the citing environment. The citing environment is furthermore characterized by hospitals (Aarhus University Hospital, Aalborg Hospital, Tweesteden hospital and General hospital de Tjongerschans, most frequently) and a mix of professional organizations, as well as by the ministry of Health and by governmental research organizations like the Cochrane Centre. Rough estimations of University and non-university groups using counts on keywords in the addresses result in approx 72% of the citing publications having a university address and 28% are (co-) published by a non-university institution, (frequently hospitals).

5   Repp

For each of the three social domains an equal number of benchmarks (five) are combined to construct a simple graph that shows involvement and activity in each of these domains. Thus, in total 15 benchmarks represent a wide set of information on the variegated work of the group. The Research Embedment and Performance Profile (REPP) is given below in the form of a simple table. In this,

— means a benchmark score of below approx. 50% of expected level

– means a benchmark score of between approx. 50% and 75% of expected level

= means a benchmark score of between approx. 75% and 100%

+ means a benchmark score of between 100% and 125%

++ means a benchmark score of higher than 125%

Table 4. Research Embedment and Performance Profile – REPP

Science, certified knowledge	
relative citation impact	—
productivity: scientific publications	++**[1]
international visibility and collaborations	=
representation in editorial boards	++
invited lectures	++
Industry, market	
non-academic/commercial citing environment	++
productivity professional publications	++*[1]
involvement in industry/market	–
advisory and expert roles in commercial domain	—
editorships professional journal	++**
Policy, societal	
involvement in policy domain	+
memberships and expert roles in governmental bodies	++
memberships of societal organizations: advisory/education	++*
production of public goods	+
additional grants societal/policy	+

*   Exceptionally high.
**   Productivity International Publications only.
[1]   note the very high percentage of senior staff (more than double that of the average program). Estimated correction for that would bring the productivity at an average level for the scientific publications (but note restriction to international publications), but still at the ++ level for professional publications.

The Repp does not show a consistent pattern in the domains, giving results that corroborate involvement in each of the domains. Remarkable is the high production of the group both in the international scientific journals and in the professional journals, but we have to note the high percentage of senior staff in this program. Furthermore the good representation in editorial boards is noteworthy. Visibility and impact in the international scientific domain are relatively modest. This, however, might be due to the specific positions of this field of research between the medical and the pharmaceutical fields. The program is not involved in commercial activities and is only somewhat concerned with industry. In the Repp the high involvement in professional journals stands out in this domain, next to visibility in this domain (hospitals). In the policy, societal domain the number of expert and advisory roles are numerous: monitoring, outcome reports, regulating issues, safety profiles and education are examples of the functions of this group in this domain. The program performs projects in collaboration with the Ministry of Health and professional organizations and hospitals. Reports on research are also most frequently given for governments or governmental labs and societal groups.

6 Stakeholder environment

6.1 Chart of the environment

In the environment of the group we distinguished three main stakeholder categories, (other) academic partners, industry and society. Obviously, all categories might be further refined into sub categories such as big pharma, small industry, non pharma industry, start ups, or as government, non-governmental organizations, patient groups, and professional groups. As a rule, we will discuss the environment in terms of the three main stakeholder categories, but sometimes we will make an exception. First we will provide the list of main stakeholders:

Table 5. Main stakeholders in three social domains

Academic	Industry	Government/society
Medical faculty Groningen - Prevend	Glaxo/Quality Institute Health Care (Kampen)	Ministry of Health, The Netherlands -Formularia and guidelines
GUIDE	Jansen Pharmaceutica	College van Zorgverzekeraars BOG
Faculty WFN	ABBOTT bv	Stichting Health Base
National Library of Medicine USA	Centrifuge	Science shops clients
University Groningen, Dept. of Biological pharmacology		InterActie apothekers; Pharmaceutical practice
University Groningen, Dept. of Clinical pharmacology		GGD Amsterdam
University of Aarhus – BIOMED		HIV group Rotterdam – Pharmacoeconomy
University of Kopenhagen – BIOMED		Psychiatry Academic Hospital Groningen and in Assen
		Child psychiatry
		Ministry of Public Health Netherlands – APODAT
		EUROCAT
		Winap/KNMP

The list shows that the group is well embedded in its societal environment, in particular in the Netherlands. Academic collaboration is a mixture of national and foreign (mostly European) partners. This reflects the group's main goal, which is to bridge the gap between pharmaceutical research and societal practice. This happens not only by practicing applied research after specific questions, but also actively helping develop better practices (protocols, use of drugs).

6.2   Stakeholder survey

We sent questionnaires to six different stakeholders from all three social domains. They included:
Ministry of Public Health
WINAp/KNMP
College voor Zorgverzekeraars
InterActie Apothekers
Stichting Health Base
Glaxo/Quality Institute Pharmacy

Of these, three stakeholders replied. All three can be classified within the pharmaceutical practice, clearly the strongest link of this group with the environment. Yet, all three stakeholders primarily

collaborate with the group with a scientific goal in mind: co-publication of articles. All three see themselves primarily as scientific colleagues and two of the three also play the role of customer and intermediary. Two of the three also collaborate to exchange people, and the third is working on a joint product (exams for pharmacists). The reasons why these stakeholders are collaborating are manifold. They want both to stimulate fundamental research and to help diffuse the results of research. All three stakeholders mention informal contacts as a way to influence the research priorities of the group. Two of the three also participate in more formal settings such as steering committees.

Although these stakeholders are generally positive about the collaboration, they also mention things to be improved: more room for truly scientific collaboration (co-publications) and more room for people exchange (two of the three stakeholders). They praise the easy accessibility of the group, the good contacts with pharmaceutical practice, and its scientific level. An important element in the kind of research this group conducts is the personal contacts the researchers maintain with the stakeholder environment. According to the stakeholder survey, contacts with pharmacists are quite positive and the general response is that they are in close co-operation with the group. Stakeholders indicate that the chances for both informal contacts and the options for more formal deliberations are quite possible. Many co-operations result in a joint publication. As a weakness a lack of coherence was mentioned (by one stakeholder), and the fact that there are no permanent staff positions in the pharmaceutical practice.

6.3 Societal position of the group

This is a young group in its building phase. Doing fundamental research is always tied to a societal practice or question. The group puts much effort in communicating results to society (and policy). It uses all kinds of media: newspapers, radio, tv, and increasingly the internet. It consults users (patients) on a regular basis. Furthermore, it stimulates people from practice (pharmacists) to participate actively in research (transdisciplinary research). When it collaborates with industry, its goal is to find a balance between commercial and scientific interests. Arguably, the pharmaceutical practice is for this group what the experimental lab is for other research groups. The researchers are very involved in the professional sector and engaged in use of drugs by patients. Their publication strategy is therefore also oriented towards national (scientific) media and professional journals.

The societal impact of the group's research is not only present in the 'practical' work (e.g. the 'InterActie' database) but also in the actions that are undertaken by the group to influence the policy agenda. Examples of this are expressions of concern on the status of the use of folic acid during pregnancy, use of medicine by children, and a permanent attention for orphan drugs. A policy-relevant research project of another kind deals with the reports by the group on cost-effects of medicines. Furthermore, the group is represented in several working groups of the Dutch Health

Council and the CvZ. Typically for this group is the goal of agenda setting. Many advisory positions function not so much to influence the national scientific research agenda, as well as to initiate and stimulate research that is relevant for policy and the professional practice of pharmacists.

The kinds of users represented in the stakeholder chart indicate that the impact of the group's research is closely directed to end-users. Where professional alliances are manifest in the close co-operation with pharmacists, the group has a direct link to end-users as well in the presence of the 'science shop'. Moreover, an explicit goal of the group seems to be the contact with a general audience, the 'public', as is manifest in publications as 'Wat doe ik, slik ik medicijnen' (edition 500.000).

7   Feedback on mission statement

Whereas many pharmaceutical research groups do have the laboratory as their home base, the main area for this group seems to be 'practice'. The group's focus is at the end of the R&D trajectory where drug utilization actually takes place. This means that the group is in constant interaction and participation with pharmacists to connect research and practice. The work on the 'Interaction' database is a good example of that kind of research, as is the project on Pharmacy Practice, although there does not seem to be a permanent staff position for the latter.
The implication of this orientation is that much of the output of the research by way of publications is directed towards the national domain of academics and professionals. While on the one hand this results in the loss of a more international academic profile, it gains academic support on the national level for innovations in the practical work of professionals.

Appendix 3
Benchmarks agricultural sciences

Explanation of benchmarks and indicators of the REPP in Agricultural Sciences study

(5 domains: collaboration and visibility; science and certified knowledge; education and training; innovation and professionals; public policy and societal issues policy)

Collaboration and visibility

Indicator	Data set	Expression in %	Meaning in 100%
Percentage personnel leaving for job in KCW	9	% mobility to KCW	100% = all personnel leaving finds next job in KCW
Percentage of projects collaborating with or funded by KCW programs	3	KCW coop./fin. proj.	100% = all projects
Percentage co-publishing KCW programs from all co-publishing institutes	1	KCW co-pub.	100 % = only co-publishing with KCW programs
Percentage Dutch co-publishing institutes from all co-publishing institutes	1	% co-publ. Neth. (not KCW)	100 % = only co-publishing with Dutch programs
Percentage Dutch citing institutes from all citing institutes	2	% Neth. citing (not KCW)	100 % = only cited by Dutch programs
Percentage foreign co-publishing institutes from all co-publishing institutes	1	% co-pub. internat.	100 % = only co-publishing with foreign programs
Percentage foreign citing institutes from all citing institutes	2	% Internat. citing	100% = only cited by foreign programs

Science and certified knowledge

Indicator	Data set	Expression in %	Meaning in 100%
All cited journal articles[1]/all journal articles	2 and 6	% cited journ. art.	100 % = all journal articles at least cited once
Total citations to journal articles[2]/ (journal articles X 2)	2 and 6	2 cit. / journ. art	average 2 citations per journal article
(journal articles X 2)/ total fte scientific personnel[3])	6 and 5	2 journ. art./fte total	average 2 journal articles per fte
Memberships editorial boards/(fte 'WP senior 1997'[4] X 1,5)	8 and 5	1,5 member sci. journal/ fte WP	average 1,5 member per fte WP other
Percentage as reported by program	7	Orientation on science (in %)	100% = total orientation on science
Second money flow (in fte's) as percentage of total fte	5	Second fte (20% of fte total)	20% of total input in fte's
Percentage of projects collaborating with or funded by a research program (including KCW programs)	3	% coop./fin. proj. →res. Grps. (incl.KCW)	100% = all projects
Percentage personnel leaving for job in research[5] (including KCW)	9	% mobility to research (incl. KCW)	100% = all personnel leaving finds next job in research

Education and training

Indicator	Data set	Expression in %	Meaning in 100%
Total dissertations (category 1)/(total fte aio+oio/4)	6 and 5	Diss. (cat 1)/4 fte aio=oio	average of 1 diss. per 4 fte Ph.D.[6]
Total fte Junior research staff/(total fte senior research staff[4]/2) as reported in 1997	5	Junior, aio, oio students (1 Jun/2 Sen 1997)	average of 1 fte junior per 2 fte senior
Total dissertations (category 1)/total persons junior staff excluding those started in 1996 and 1997[7]	6 and 5	# diss. (cat. 1)/ # junior staff	1 dissertation per junior program member started before 1996[8]

Innovation and professionals

Indicator	Data set	Expression in %	Meaning in 100%
Percentage personnel leaving for job in firm[5]	9	% mobility to company	100% = all personnel leaving finds next job in firms
Third money flow (in fte's) as percentage of total fte	5	Third budget fte (45% of fte total)	45% of total input in fte's
Percentage as reported by program	7	Orientation on professionals (in %)	100% = total orientation on professional domain
Percentage of projects collaborating with or funded by commercial or innovative organization	3	Coop./fin. proj. innovating inst.	100% = all projects
Number of patents or royalty contracts/(total fte 'WP other'/3)	10	1 patent/3 WP	average of 1 patent or royalty contract per 3 WP-other
(professional articles X 2)/total fte scientific personnel[3]	6 and 5	2 prof art/fte total	average 2 professional articles per fte
Memberships (scientific) advisory boards/(fte 'WP senior 1997'[4] X 1,5)	8 and 5	1,5 member advisory board/WP	average 1,5 member per fte WP other

Public policy and societal issues

Indicator	Data set	Expression in %	Meaning in 100%
Percentage of projects collaborating with or funded by special programs, government or ngo's	3	% coop./fin. proj. gov./spec. prog.	100% = all projects
Memberships advisory committees governmen /(fte 'WP senior 1997'[4] X 1,5)	8 and 5	1,5 members gov. or spec. prog./WP	average 1,5 member per fte WP other
Percentage personnel leaving for job in policy arena arena	9	% mobility to gov./policy	100% = all personnel leaving finds next job in policy
Estimation of involvement, each relationadds 5%	11	Involvement NGOs (scored in %)	100% approx. 20 relations

Notes

1. Refereed journal articles in 1993-1994-1995-1996 cited at least once in 1994-1995-1996-1997.
2. All citations in period 1994-1995-1996-1997 to journals articles published in 1993-1994-1995-1996.
3. Period 1993-1997.
4. Professor/associate professor + other senior staff + post doctoral fellows as reported for 1997.

5. Retired personnel excluded.
6. Exceeding production is indication for production of Ph.D. students next to AIO's and OIO's.
7. Source: Listing program members (question 3: VSNU self reports).
8. Although one may not expect Ph.D. students started in 1995 to have finished there dissertation in 1997 one also has to reckon with students started before 1993 finishing there thesis in the period 93-97.

Data used in REPP

(Social) Science Citation Index
1. Selection SCI/SSCI publications 1993-1994-1995-1996-1997. Assigned to research programs using team members and previous departmental address of research program.
2. Articles in SCI citing data set 6 (1994-1995-1996-1997).

WAU annual reports
3. Research project listings (titles, funding, cooperation, diffusion results) 1993-1994-1995-1996-(1997). Corrected, completed and commented on by program in survey.

VSNU self evaluation reports
4. Project members 1993-1994-1995-1996-1997 as reported by programs.
5. Input in full time equivalents.
6. Publication listings (excluding proceedings and abstracts) 1993-1994-1995-1996. Assigned to programs by programs themselves.

Survey
7. Strategic orientation indicated by program.
8. Memberships and advisory functions of: scientific organizations; societal organizations; government; companies and consultancy firms.
9. Mobility of personnel leaving research program (estimation of number of members leaving program, next organization, general description of first function).
10. Number of patents or royalty contracts.
11.

Other
11. Qualitative estimation of involvement in NGOs using project listing and response on several questions in survey (Each relation/ advisory function /cooperation adds 5%).

Appendix 4
Benchmarks pharmaceutical sciences

Explanation of benchmarks and indicators of the REPP in the Pharmaceutical Sciences study

(3 domains: science and certified knowledge; industry and market; policy and society)

Science, certified knowledge

Relative citation impact
Combined benchmark
a.   impact cited publications is 112% of impact journals and
b.   average citations of cited publications are 150% of citations to citing publications

Norm relative impact a
Observatorium: Average impact Dutch Universities relative to pharmacology journals = 1.02
KUOZ figures: On average this should be approx. 12% above world average impact for pharmacology

Norm relative impact b
No comparable figures available in bibliometric or scientometric analysis. Most programs score above these benchmarks

Algorithm
relative impact a
A = Sum of the average numbers of citations per year within two years window (1st and 2nd year after publication) to all publications with in selected set (15 publications) for which two years window is available (published before 2000),
B= Sum of impact factors[63] of all journals in which the articles used in A are published.
Relative impact a = A/B.
relative impact b
A= Number of citations to publications in selected set published before 2000,
B= the number of publications in selected set published before 2000

---

63   Impact factor Journal: Journal Citation Reports 2000. The impact factor is a measure of the frequency with which the "average article" in a journal has been cited in a particular year. **Calculation for journal impact factor. A**= total cites in 2000 **B**= 2000 cites to articles published in 1999-98 (this is a subset of A), **C**= number of articles published in 1990-91, **D**= B/C = 1992 impact factor **(NOTE: 1st and 2nd year citations are counted)**

C= All citations to publications in citing set published before 2000 (publications citing the selected set),
D= the number of publications in the citing set published before 2000
Relative impact b= (A/B) / (C/D).

Citations are counted in the Science Citation Index, the impact factor is found in the Journal Citation Reports

Productivity scientific publications
Benchmark
2,5 publications / 1 fte WP (first + second + third flow)

Norm
The average production of scientific publications differ strongly from sector to sector, ranging from approx 2,5 pub/fte to over 6 pub/fte . Definitions of scientific publications play a role in the differences between fields (KUOZ 2000).
Pharmacology = (KUOZ p31) input 192 fte output 605 scientific publ = approx. 3 sci pub/fte WP (1998). The programs assessed in this report however include other fields.
Based on external advice and comparison with the average figures of all groups the benchmark norm was set to 2,5 scientific publications for each fte scientific personnel input. The figures have been calculated for the whole period (1996-2000).
On average the programs in this study scored 2,8 scientific publication/ fte WP

Algorithm
A= the number of scientific publications published by the program in the years 1996-2001 as reported in scientific reports,
B= the total input in full time equivalents (first, second and third money flow) of the program in the years 1996-2001 as reported by the program.
productivity scientific publications = A/B

International visibility and collaborations
Combined benchmark
a.   International collaborations: 20% of the selected 15 publications of the program carry also a foreign address.
b.   International visibility: 80% of the citing publications carry a foreign address. Benchmark for a) international collaborations is partly derived from the Observatorium which gives 17% on average for all Dutch publications and 13 to 15% for all Dutch Pharmacology publications.

UIPS defines a target of 25% of all scientific publications (Strategisch Plan (Dutch edition) p28). Furthermore the benchmark is chosen to differentiate between the programs involved in this report. The average percentage of articles carrying a foreign address for the programs in this study is 29%.

The benchmark for b) international visibility is rather new, to our knowledge no comparable figures are available. The benchmark was therefore set to differentiate between programs involved. For the average program in this study the percentage of citing articles carrying a foreign address is 82%.

Algorithm
a. Percentage of publications in selected set (15 articles) carrying a foreign address
b. Percentage of citing publication carrying a foreign address (the complete address information is available in the bibliographic descriptions of the publications derived from 'Web of Science')

Representation in editorial boards
Benchmark

1,6 editorship scientific journal or scientific series for each full time equivalent in the first money flow (2001 only).

The benchmark is set to differentiate between the programs involved in this study and is without reference to other fields or comparable other groups. Differences are large between the programs, the average number of editorships varies from 0,3 till nearly 5 per fte WP 1. The average is approx. 1,6 editorship/fte P 1. The number of editorships is reported by the programs by means of the questionnaire (question 4.2.1). There might be differences in the classification and counting of the editorships between programs.

Algorithm

A= number of editorships of scientific journals an scientific series as reported by the program in the questionnaire (4.2.1.) for 2001.
B= the input in full time equivalents WP 1 as reported in scientific reports for 2001.
Representation in editorial boards = A/B

Invited lectures
Benchmark

1,5 invited lecture for each full time equivalent in the first money flow (2001 only)
UIPS sets as a target 2 invited lectures each year for all professors, WP 1 consist not only of professors but also of junior staff and assistant professors etc. The benchmark was therefore set to 1,5.

Algorithm:

A= the number of invited lectures in 2001 as reported by the program (additional information to the questionnaires),

B = the input in full time equivalents WP 1 as reported for 2001

invited lectures = A/B

Industry, market

Non-academic / commercial citing environment

Benchmark

20% of all addresses in both cited (selected) set and citing set are commercial/non university addresses as defined below.

There is no previous comparable analysis of the citing environment known to us. The benchmark is set to differentiate between to groups in this study. On average the groups score is 21%, ranging from 12% to 33%.

Algorithm

A= estimation of the number of industrial / commercial addresses in selected set (collaborations) by visual inspection

B= estimation of commercial / company addresses in citing set, by search on keywords and visual inspection

C= estimation of non-academic addresses in citing set by search on keywords and visual inspection

D= number of addresses in cited set

E= number of addresses in citing set

non academic/commercial citing environment is: $(A+B+C)/(D+(Ex2))x100\%$

The complete address information is available in the bibliographic descriptions of the publications in "Web of Science".

Productivity professional publications

Benchmark

The benchmark is chosen at 0,6 professional publication for each full time equivalent input (all money flows).

For Pharmacology the following figures are known: Input 192 fte, output 152 prof publ (KUOZ p31) on average this is approx 0,74 prof. pub/fte. We have to take into consideration that professional publications is a meaningless category for some programs for other programs professional publications are a substantial portion of there output. The average for all programs

in this report is 0,54 prof publ / fte WP (all money flows) ranging from zero to nearly two prof publications / fte WP.

Algorithm
A= the number of professional publications as reported in the scientific reports 1996-2001
B=input in full time equivalents first, second and third money flow as reported by institute 1996-2001
productivity professional publications = A / B

Involvement in industry/market
Eight different activities are distinguished as indications of involvement in the industry/market domain:
1. industry as one of main partners
2. commercial activities, like spin-off activities and start ups
3. research development /product design
4. patent applications; royalties and developing patent
5. additional grants from industry/companies
6. reporting to industry/ presentations & publications
7. influence industry on research agenda setting
8. use of research by industry

Benchmark: six out of eight activities
The benchmark is the result of an informed estimation, it results in a good differentiation between the programs of which four score below the benchmark, two just above and the rest considerably above the benchmark

Algorithm
Aactivities are counted (sometimes weighted for high or low involvement)

Advisory and expert roles in commercial domain
The following categories are reported there: Advisory functions or member of a board in: pharmaceutical industry; medical or other consultancy offices; own firm or other commercial functions.

Benchmark
1 advisory function of expert role for each full time equivalent WP first money flow.
This benchmark is also the result of an informed estimation, is also is near the average for all

programs, the range of scores however is very large, ranging from zero to nearly 4.5 advisory roles per fte WP.

Algorithm
A= sum of all advisory roles and board memberships as reported under question 4.2.4. for 2001
B= total of input in scientific personnel WP for 2001
Advisory and expert roles in commercial domain = A/B

Editorships professional journals
Benchmark
0,3 editorship professional journal, popular scientific journal or popular scientific series for each full time equivalent in the first money flow (2001 only).
The benchmark is set to differentiate between the programs involved in this study and is without reference to other fields or comparable other groups. The number of editorships is reported by the programs by means of the questionnaire. There might be differences in the classification and counting of the editorships between programs.

Algorithm
A= number of editorships of professional journal or popular scientific journals or series as reported by the program in the questionnaire for 2001. B= the input in full time equivalents WP 1 as reported in scientific reports for 2001.
editorships professional journals = A/B

Policy, societal

Involvement in policy domain
Nine different activities are distinguished as indications of involvement in the industry/market domain:
1. additional grants and partners
2. important societal partner/project
3. influence societal groups on research agenda
4. influence policy groups on research agenda
5. use of research by both government and societal groups
6. involvement in regulating/law
7. societal/public debate by public media
8. presentations and publications for government/gov. labs/societal groups
9. collaborations: societal as one of main partners

Benchmark

Seven out of nine activities

The benchmark is the result of an informed estimation, the benchmark seems to be rather high for these programs of which ten score below the benchmark, and the five above the benchmark.

Algorithm

Activities are counted (sometimes weighted for high or low involvement)

Memberships and expert roles in governmental bodies

Benchmark

0,7 advisory roles (as defined below) for each full time equivalent input scientific personnel first money flow (2001 only).

The benchmark represents the average score of the programs involved in this study, there is a large difference between the programs ranging form approx. 3,5 memberships to zero memberships.

Algorithm

A= the sum of all relations with government as reported in the questionnaire comprising: advise to parliament, memberships of governmental advisory bodies; memberships of departmental committees of advisory bodies; memberships of international; national or regional committees or advisory bodies (2001).

B= Input in full time equivalents first money flow 2001 as reported in scientific reports.

memberships and expert roles in governmental bodies = A/B

Memberships of societal organisations: advisory/ education

Benchmark

1,1 advisory roles (as defined below) for each full time equivalent input scientific personnel first money flow (2001 only).

The benchmark represents the average score of the programs in this study, there is a large difference between the programs ranging from approx. 6,5 advisory roles to zero memberships.

Algorithm

A= the sum of all relations with government as reported in the questionnaire comprising: membership of advisory body (societal); educational activities in this domain and memberships of interest groups (2001).

B= Input in full time equivalents first money flow 2001 as reported in scientific reports.memberships of societal organisations: advisory/education= A/B

Production of public goods

Benchmark
Reporting five examples
This indicator is only a rough estimation, it distinguishes between the programs convincingly involved in the production of public goods and those less clearly involved.
A= number of examples of product with societal value as reported in the questionnaire under q3.9.
production of public goods = A

Additional grants from policy

Benchmark
27% of the six mentioned options for additional financing are funded by societal/governmental organisations.
This benchmark is an estimation to differentiate between the programs in this exercise.

Algorithm
A: number of items mentioned for with additional grants are received by societal or governmental groups of from mission oriented funds. The issues are: Instruments and equipment; bench fees and consumables; technical personnel; workshop; other goals. (financing of scientific personnel is excluded in this indicator).

Appendix 5
Action plan Evaluating Research in Context

The project "Evaluating Research in Context" (ERiC – www.eric-project.nl) originally emerged from a COS-project about the societal quality of research. Under the aegis of that project the sci_Quest method as discussed in this book has been developed. The method was received quite well in the Dutch academic World, and appreciated by the review committees that worked with it. Following the publication of the first edition of *Evaluating Research in Context* in 2005, it was therefore decided to establish a broader group in which NWO, KNAW, VSNU, QANU and, more recently, the HBO-Raad participate. These organisations considered the time ripe for a joint effort regarding the development of new methods for evaluation that included the societal quality of research. The cooperative group is referred to as Context Group (CG).
The CG developed an action programme for 2006 and 2007 responding to the needs enunciated by the minister of education and sciences as well as by research institutions.

The action plan, Evaluating Research in Context, is directed toward two primary goals:
1. the advancement of both sharing knowledge and raising consciousness in the Dutch academic world, in particular among university administrators and scientific staff who regard the introduction of societal impact indicators in the national evaluation system.
2. the stimulation of international knowledge exchange and methodological development in this field.

It includes the following work packages:
» Quick scan of the field of international evaluation methods for the societal quality of research.
» Practical guide or handbook on evaluation of societal quality and relevance of research.
» Workshops for institutions (researchers and policy makers) that want to consider societal criteria in their evaluation procedures, through the use of the sci_Quest or other methods.
» International expert seminar (9 November 2007) on the evaluation of scientific research in societal context.
» Study into the need for a support office for evaluations of the societal quality of research, that can provide support to institutions that want to evaluate their research in a broader manner.
» Second edition of the book *Evaluating Research in Context* with a new introduction.

The Context Group consists of the following persons who also serve as contacts for the various organisations.

Henriëtte Maassen van den Brink	chairwoman
Leonie van Drooge	executive secretary
Pierre Morin	COS
Jack Spaapen	KNAW
Jacco van den Heuvel	KNAW
Renee Westenbrink	VSNU
Victor van Rij	COS
Frank Wamelink	QANU
Frank Zuijdam	NWO
Johannes van de Vos	HBO-Raad

Observers

Jan van Steen	Ministry of Education, Culture and Sciences
Jacqueline Hulst	Hogeschool Utrecht

The Context Group is supported by an international group referred to as the Resource Group. It's present membership includes the following persons.

Name	Affiliation
Erik Arnold	Technopolis UK
Julieta Barrenechea	UPV/EHU, San Sebastian
Peter Groenewegen	NIVEL, Utrecht
Peter van den Besselaar	Rathenau Institute, The Hague
Wim Blockmans	Netherlands Institute of Advanced Science, Wassenaar
Wiebe Bijker	University of Maastricht
Luke Georghiou, John Rigby	PREST, MBS, Manchester
Ken Guy	Wise guys LTD
Andre Knottnerus	University of Maastricht, Gezondheidsraad
Stefan Kuhlmann	Fraunhofer Institut, University of Twente
Philippe Laredo	École Nationale des Ponts et Chaussées
Catherine Lyall	ESRC Edinburgh
Barend van der Meulen	University of Twente
Jordi Molas Gallart	INGENIO (CSIC), Universidad Politécnica de Valencia
Helga Nowotny	ETH Zürich, European Research Advisory Board

Wija Oortwijn   Rand-Europe

Julie Thompson Klein   Wayne State University Detroit

Terry Shinn   CNRS, Maison de Science de l'homme

Peter Weingart   Universität Bielefeld

Appendix 6
Abbreviations

AIO	Research Assistant
COS	Consultative Committee of Sector Councils for research and development
DLO	Center for Agro-biological Research
FTE	Full Time Equivalent
GRIP	Groningen Research Institute of Pharmacy
KCW	Wageningen University Research Center
KNAW	Royal Netherlands Academy of Arts and Sciences
KUOZ	Index Numbers University Research
LUW	Wageningen University
MIT	Multi-, Inter- and/or Transdisciplinary Research
MKO	Societal Quality of Research
NGO	Non Governmental Organization
NRLO	National Council for Agricultural research
NWO	Netherlands Organisation for Scientific Research
OECD	Organisation for Economic Co-operation and Development
OIO	Trainee Research Worker
QANU	Quality Assurance Netherlands Universities
REPP	Research Embedment and Performance Profile
RUG	University of Groningen
SCI	Science Citation Index
SSCI	Social Science Citation Index
SEP	Standard Evaluation Protocol
SWR	Social Sciences Council
UIPS	Utrecht Institute for Pharmaceutical Sciences
UU	Utrecht University
VF	Conditional Financing
VSNU	Association of Universities in the Netherlands
WP	Scientific Personnel

COS HBO-raad

The Consultative Committee of Sector Councils for research and development (COS) in the Netherlands is the collaborative platform for sectorcouncils and other members specialised in foresight studies.

Functions of the COS are e.g. promoting a joint approach in foresight- and programming studies as well as studies on the development of methodology, funded by the COS Coordination Fund. Furthermore the COS sees to joint input in administrative consultations with ministries and other organisations. The COS is a member of the European Research Area (ERA) Network Forsociety in which 15 EU member states are working together on joint foresight activities en benchmarking.

The most important function of the sector councils and the COS as a cooperative platform, is providing a space in which representatives from government, society, business and research can reflect on and discuss social and technological developments from a long-term perspective (10-20 years) – reflections that can then be used to inspire knowledge and policy agendas.

The COS members endeavour to gain a good sense of the problems, possible solutions and demand for knowledge that the future will bring, while simultaneously keeping abreast of newly developing scientific insights and technologies. The results of their efforts are used by society and industry, governments, NGOs, the Dutch Organization for Scientific Research (NWO) and universities to support the development of innovative policies and research.

Participatory, societally oriented foresight is an important instrument. It creates a multifaceted images of problems, developments and new chances. Bringing diverse groups together also has the advantage of tapping into greater amounts of creativity and forming new networks that often play an important role in setting up research- and innovation oriented programmes and in realising policy strategies.

In 2005 the COS began carrying out bi-annual "horizon scans" in order to select important themes for its foresight activities. The horizon scans are directed toward making visible the future's most important problems and threats, as well as promising new developments (opportunities) for solving problems and furthering human progress.

For more information please visit **www.minocw.nl/cos**.

KNAW
Royal Netherlands Academy of Arts and Sciences

The Royal Netherlands Academy of Arts and Sciences (Koninklijke Nederlandse Akademie van Wetenschappen) was founded in 1808 to promote learning in the Netherlands. Active scientists with an excellent scientific record are eligible for membership of the Academy, which is for life. The Academy has two divisions: the Science Division (mathematics and physics, life sciences and technical sciences), with 124 members, and the Humanities and Social Sciences Division (humanities, law, behavioural sciences and social sciences), with 102 members.

Mission

As the forum, conscience, and voice of the arts and sciences in the Netherlands, the Royal Netherlands Academy of Arts and Sciences promotes the quality of scientific and scholarly work and strives to ensure that Dutch scholars and scientists make the best possible contribution to the cultural, social, and economic development of Dutch society.

Main functions

» *Advising the government on matters related to scientific research*
» The Academy is supported in its advisory capacity by councils for different scientific fields. These bodies are composed of both members and non-members of the Academy, including university professors, and scientists from public and private research institutes as well as industrial laboratories. Counsel, be it solicited or not, is given to: the government, the universities, research organisations, funding agencies and international organisations.
» *Assessing the quality of scientific research (peer review)*
» The Academy makes available its expertise for assessing the quality (by peer review) of basic and strategic research performed in institutes, research groups or within the framework of specific programmes. It awards fellowships to eminent professors who are at an advanced stage of their scientific career. The Academy is responsible for accrediting university institutes for postgraduate research training.
» *Providing a forum for the scientific world and promoting international scientific cooperation*

» The Academy actively promotes science in general and facilitates exchange and co-operation between scientists at home and abroad. It organises and funds conferences and symposia, collaborates with sister academies and enables Dutch scientists to participate in conferences elsewhere.
» *Acting as an umbrella organisation for the institutes primarily engaged in basic and strategic scientific research and disseminating information*
» The research institutes of the Academy carry out basic and strategic research in the life sciences, humanities and social sciences. Some of the institutes also have a scientific service function by forming and managing biological and documentary collections, providing information services and creating other facilities for research. The Academy's institutes, which are located throughout the country, employ a total of approximately 1.300 staff.

NWO
The Netherlands Organisation for Scientific Research

The Netherlands Organisation for Scientific Research (NWO) finances 5600 excellent researchers working at universities and institutes and stage-manages Dutch science.

Mission and ambition

The NWO has the following statutory mission. The Netherlands Organisation for Scientific Research:

» is responsible for enhancing the quality and innovative nature of scientific research as equally initiating and stimulating new developments in scientific research
» mainly fulfils its task by allocating resources
» facilitates, for the benefit of society, the dissemination of knowledge from the results of research that it has initiated and stimulated
» mainly focuses on university research in performing its task.

In fulfilling its responsibilities NWO pays due attention to the aspect of coordination and facilitates this where necessary. NWO wants to ensure that Dutch science continues to be amongst the best in the world and that the currently strong position is further strengthened. NWO would also like to see a more intensive use of the results from scientific research by society, so that the contribution of scientific research to prosperity and welfare can be further increased.

VSNU
Association of Universities in the Netherlands

All Dutch research intensive universities, 14 in number, are member of the Association of Universities in the Netherlands (Vereniging van Universiteiten, VSNU).

The VSNU advocates the interests of the research universities regarding research, education, knowledge transfer, funding and personnel policy. On behalf of the universities, the VSNU signs the Collective Employment Agreement (CAO). The VSNU is a member of the European University Association (EUA).

QANU

Quality Assurance Netherlands Universities (QANU) offers universities external assessments of academic education and research programmes, and advice on ways of improving internal quality assurance. QANU's services include:

» peer review of university education and research
» support for submission of applications for accreditation from universities in the Netherlands and abroad
» advice on improvement of internal quality assurance

QANU works independently of universities, within the statutory framework set up for the assessment, accreditation and funding of university education and research in the Netherlands.

QANU's key activity is the external assessment of academic bachelor's and master's degree programmes in Dutch universities. QANU has drawn up an assessment protocol permitting transparent, systematic and reliable assessment of the programmes in question against both national and international benchmarks. The writing of a self-evaluation report by the course provider, and site visits by an external panel, form crucial parts of the assessment procedure. This panel writes an assessment report which the university must submit to the NVAO together with its application for renewal of its accreditation.

QANU is also offering external assessment of research programs in accordance with the Standard Evaluation Protocol 2003–2009 for Public Research Organisations formulated by the KNAW (Royal Netherlands Academy of Arts and Sciences), the NWO (Netherlands Organisation for Scientific Research) and the VSNU (Association of Dutch Universities) stipulates that universities must carry out a self-evaluation of their research activities once every three years, and that these research activities must also be assessed by an external panel once every six years. The external assessment covers not only the content of the research programme but also the management, strategy and mission of the research centre where it is carried out. QANU underwrites the importance of methods for the assessment of societal quality of research in certain areas of scientific activity.

For more information please visit: **www.qanu.nl**

About the authors

Jack Spaapen is coordinator Quality Assurance and Evaluation at the Royal Netherlands Academy of Arts and Sciences (KNAW). His sci_Quest work involves the areas of research evaluation, research policy and science and technology questions related to development countries. He can be reached at jbspaa@xs4all.nl.

Huub Dijstelbloem is working for the Rathenau Institute as coordinator Technology Assessment. He published about the beginnings of Aids in the Netherlands, the role of the Internet in public information services, environmental policymaking in the EU and about John Dewey's conception of democracy. He can be reached at dijstelbloem@gmail.com.

Frank Wamelink is coordinator degree programme assessments of Quality Assurance Netherlands Universities (QANU). His main task is to manage the assessments by independent peer review committees of the degree programmes delivered by the Dutch research based universities to attain mandatory accreditation. He has been involved in sci_Quest projects from the start. He is especially interested in the combination of both qualitative and quantitative evaluation methods, the involvement of stakeholders in assessments and the incorporation of several forms of societal impact. He can be reached at f.wamelink@planet.nl.

The sci_Quest Research Team

The sci_Quest Reseach team responsible for the empirical work in the two studies in agricultural sciences and pharmaceutical sciences consisted of the following persons:

Senior researchers:
Huub Dijstelbloem
Jack Spaapen
Frank Wamelink

Junior researchers:
Martin Boeckhout, student in the department of Science Dynamics of the University of Amsterdam
Janelle Ward, student International School of the University of Amsterdam

Consultants:
Jet Bout, Conscience, Amsterdam
Ad Prins, PhD in Science studies, independent scholar (Groningen)Olor senisiscil ullum inci blam, summy num do od modip eu feugiat, qui estrud enisi bla cortinc iduiscin vulputatis am velit acilla faciduis exeros nullam, vercilit augue dio delenit lumsandio dunt aliquat, consequisl dipsum dolorem dolore feum vero odit vullummy nim iniam ex etue ecte conum eum veliquip et wisi. Feumsan dreetum dolobore feu facilit nullamet vulla faciliscil eu faccum in essi blan ulpute conse venim in er sismolorem amet acilit la facinci tatet velit luptat, consecte dunt diamet, quat. Henis nonsenisi bla feu feum quat. Dui te dolor sumsandreet lut er si ecte velessi tie feu feugiam consenit wis euis atue vel ipit, volore faccum zzrit num quat. Ut aciduis augait nos am, vent nos aciliquatio commy non veliqui psummy nit lortionum estrud tat incinis nos dolor sed tatum nim et in vulputat. Enis euisis amet, consed eu feu feugue magniat. Ut velisci et, venissim ip enit dolore consectem vel et, velesequi bla faccum dolobore feum inci ectem dio enisl iriure modiat nim alit, qui tatie faci blaorper se doloreet, volore moloborper se ea amconse quatie dunt lorpero odolore er iliqui blandre consenim venibh exer sum ip ea faccummy num alit praessit ilisi enibh et prating ex eu facilit nullutp atinit, core minisit iliquat. Et ad exero conullutat in utpat am esequipit aliquat duissit nullametuer sis nisit pratetue dit in utpatum ipisi.
Do doloreriurer aliquis ad tie conseniam, vel euguerit wis ea feu faccum zzril ipiscilisit aliquamet dio do esequam consed magna ad min velit et wis el utat. Ut in verci ecte tat. Ut wisl el dolor sum alis ad dolorpe rcidunt do dolore diat. Lum nonsequat nos nonse minis am del el et acilit aliquam zzrit, quatummy nulputet am, velenibh esed elessequipit euip ex exeriurem iriuscidunt la feugait num do euis etum dolorem vullamconsed minis erit aut lan volorem vel et atumsan hendre magna conse tatue ming enim zzril doluptat aliquat vullam dionsequate et, commodi amcommodio cons nim nisis eu feuguer aut vulluptat. Duisl dunt am, sustrud el iuscipit dolore faciduisl irit wis autet