QUALITY INDICATORS
FOR RESEARCH IN THE HUMANITIES

Royal Netherlands Academy of Arts and Sciences
PO Box 19121, NL-1000 GC Amsterdam
T +31 (0)20 551 0700
F +31 (0)20 620 4941
knaw@bureau.knaw.nl
www.knaw.nl

pdf available on www.knaw.nl

Basic design edenspiekermann, Amsterdam
Typesetting: Ellen Bouma, Alkmaar

Illustration cover: Dragan85

ISBN: 978-90-6984-629-3

The paper for this publication complies with the iso 9706 standard (1994) for permanent paper

QUALITY INDICATORS FOR RESEARCH IN THE HUMANITIES

Royal Netherlands Academy of Arts and Sciences
Interim report by the Committee on Quality Indicators in the Humanities
May 2011

FOREWORD

The scarcity of funds for scientific research and the wish to allocate those restricted funds to high-quality research make it increasingly important to measure and assess the quality of research. This also applies to the humanities.

There is, however, fairly widespread discussion of what constitute the right methods of quality assessment, including in the humanities. In 2005, the Academy published the report *Judging Research on its Merits*, in which the predominance in evaluation of methods from the natural and life sciences was criticised. It argued, amongst other things, for 'special' measuring methods for the humanities and social sciences and for the social value of these scientific disciplines to be taken into consideration in assessments. In 2009, the Committee for the National Plan for the Future of the Humanities (the 'Cohen Committee') stated in its report that in the humanities there is an 'inadequate range of tools for objectively defining and differentiating quality'. The Cohen Committee therefore recommended that the Academy should take the lead towards developing the simplest, clearest, and most effective system of quality indicators for the humanities in the Netherlands. The Academy gladly accepted this task, also in light of the views expressed in the 2005 report mentioned above.

In October of 2009, the board of the Academy installed the Committee on Quality Indicators in the Humanities with the following remit: to propose a system of quality indicators for the humanities based on an international survey of the existing methods and a consultation of the humanities sector. The effectiveness of the system should be tested through pilot studies in a number of different areas.

This interim report is an account of the findings of the Committee and includes a proposal for an evaluation system based on indicators which will now be tested in practice. The board of the Academy has noticed that in a number of important ways the system proposed by the Committee bears resemblance to what was put forward in another report recently issued by the Academy Committee for constructing and design sciences. Both reports suggest focusing evaluations on two main criteria, scientific

quality and societal relevance. Both reports recommend using indicators in three categories: output, use and recognition. And both reports maintain peer review as the core of evaluation procedures. Given these parallels, the Academy will encourage a broader discussion to see whether these common elements may lead towards a novel approach of quality evaluation in some or perhaps all evaluations. Together with the rectors of the universities and the ERiC-project, the academy initiated an inventory of recent studies regarding this question. The results will be presented shortly after the summer break of this year.

At the same time, the board of the Academy is looking forward to the results of the two pilot studies that will be conducted at the Meertens Institute of the Academy and at The Groningen Research Institute for the Study of Culture (ICOG) of the University of Groningen. It is expected that both studies (the inventory and the pilots) will profit from each other and that they will deliver building blocks for a new approach to evaluation in which the wealth of scientific research in a variety of fields can be accounted for.

Robbert Dijkgraaf
President KNAW

CONTENTS

SUMMARY

Background

In 2009, the Committee for the National Plan for the Future of the Humanities (the 'Cohen Committee') stated in its final report (*Sustainable Humanities* [*Duurzame Geesteswetenschappen*]) that in the humanities there is an 'inadequate range of tools for objectively defining and differentiating quality'.[1] The Committee made the following recommendation to the Royal Netherlands Academy of Arts and Sciences (KNAW, 'the Academy'): 'Take the lead towards developing the simplest, clearest, and most effective system of quality indicators for the humanities in the Netherlands'.[2]

The Academy adopted that recommendation and, in October 2009, set up a Committee on Quality Indicators in the Humanities with the following remit:

- to produce an international inventory of the existing methods for assessing quality in the humanities (inventory phase);
- on the basis of that survey, and after extensive consultation with the humanities sector, to design a system of quality indicators for the humanities in the Netherlands (conceptual phase);
- by arranging for pilot projects, to carry out empirical testing of the effectiveness of that system (empirical phase) and if necessary to make changes to the system on the basis of the results;
- to produce a final advisory report.

In the present report, the Committee presents its interim findings prior to the pilot phase of the project.

1 *Commissie Nationaal Plan Toekomst Geesteswetenschappen* (2008) 41.
2 *Ibid*., 45.

The Committee finds it important to note that the expectations of some policy-makers are too positive regarding a 'simple' and preferably fully quantified system that makes easy comparisons possible between research groups and even whole disciplines. On the other hand, the Committee notes that 'in the field' there is still a significant aversion to the idea of 'measuring' quality and to the imposition of what are considered to be management models taken from the business sector, and there is widespread concern regarding the increasing bureaucratic burden imposed by the culture of evaluation. The Committee views the need for evaluation and accountability as a given, but it has sought a system of quality indicators that is as effective as possible in so far as it attempts to do justice to the variety of forms and of scholarly practices in the humanities and elsewhere, but that is also practicable and that does not lead to an excessive bureaucratic burden.

The Committee bases its findings on a survey of international research on quality assessments and indicators, a survey of recent developments in other countries, and interviews with a large number of those concerned in the Dutch context.

Quality and quality indicators

In many respects, the quality of scientific/scholarly research is a relative or relational concept: it is assigned by other parties in relation to a particular usage context and on the basis of certain external standards. Depending on the usage context, it is possible to distinguish various aspects of quality. Research quality indicates the usefulness of the results of research for the scientific community within one or more disciplines. Societal quality indicates the usefulness of the results of research for a wider group: government, business and industry, cultural and civil-society institutions. The latter aspect has gained in importance with the increasing interest in knowledge valorisation in recent years. By emphatically giving societal quality a place in its reporting, the Committee links up with the report *Impact Assessment Humanities* [*Impact Assessment Geesteswetenschappen*] published by the Netherlands Organisation for Scientific Research (NWO) in 2009[3] and the *Guide for Evaluating the Societal relevance of Scientific Research* [*Handreiking Evaluatie van Maatschappelijke Relevantie van Wetenschappelijk Onderzoek*] resulting from the ERiC project involving the Academy, the NWO, and the Association of Universities in the Netherlands (VSNU) (2010).[4]

In actual practice, research quality is assigned by the researchers' peers. Peer review is broadly accepted, and the Committee too is convinced that it is an indispensable tool. Peer review can, however, be supported and reinforced by external quality indicators, i.e. all the indications of quality that can be discovered beyond the content of the research results themselves as assessed by peers. Such external quality indicators comprise both output indicators (for example publication in 'A' journals and also perhaps certain bibliometric indicators) and indicators of esteem (prizes, scholarly

3 Broek & Nijssen (2009)
4 ERiC (2009)

positions, and other evidence of external recognition). External quality indicators support the process of peer review by providing it with an intersubjective basis and by structuring it. A system of quality indicators is practicable if it is as simple and economical as possible as regards provision of the necessary information by the parties that are being assessed. Explicit attention will be paid to this aspect during the pilot phase, partly in relation to the current system of reporting in the METIS research information system.

Quality indicators in the humanities

There is no reason to adopt a different approach to the concept of research quality in the humanities to that adopted in the other fields of science and scholarship. In the humanities too, quality exists in the sense of the significance of the results of research for the scholarly community and for others. Here too, it is primarily the researchers' peers who must be considered able, with the aid of external indicators, to judge the quality of the research. The humanities therefore do not require their own particular type of quality indicators.

The humanities do, however, demand a fairly wide range of quality indicators that does justice to the diversity of products, target groups, and publishing cultures that are present within this field. Fair consideration needs to be given to monographs and international publications in languages other than English, meaning, for example, that primacy cannot be accorded to bibliometric indicators that are as yet still based on databases consisting primarily of English-language journal publications. The system must also offer scope for other types of output than just scholarly publications, for example databases, catalogues, and editions of texts. The system needs to be as broad as scholarly practice demands, but at the same time sufficiently flexible to enable a tailor-made approach in various contexts and to remain practicable. By emphasising the importance of a flexible system with the option of including context-specific indicators – related to the specific character of a given discipline or the mission of a given institute – the Committee endorses the findings of the Academy's report *Judging Research on its Merits* (2005).

Peer review

For two reasons, the report gives separate consideration to the advantages and disadvantages of peer review as a basis for quality assessment. It does so firstly because quality indicators need to operate within a system of peer review and are intended to minimise the risks associated with peer review (particularly subjectivity) and secondly because external quality indicators themselves find their basis in peer review. They are, as it were, a reflection of repeated assessment by fellow scholars. Criticism of peer review as such can therefore basically undermine the value of external indicators.

The Committee concludes that the most serious objections to peer review can be neutralised by means of procedural guarantees and by utilising external indicators that give an intersubjective basis to the judgment of peers. The Committee also concludes that the common objections to peer review only detract slightly from the usefulness of peer review-based external quality indicators 'because in assessment procedures they are ideally not utilised in isolation, but in combination with other indicators.

Bibliometric indicators

The report also considers bibliometric indicators separately. A number of objections, ranging from principles to practice, can be made to the ill-considered use of bibliometric indicators in general and to their use in the humanities in particular. As a matter of principle, there is no one-to-one relationship between impact and quality; there are major differences in citation culture between disciplines, between subjects within disciplines, and within language areas (the English-speaking or French-speaking worlds, for example) that make comparisons between citation scores difficult. There is also a major mismatch between the often lengthy 'half-life' of publications in most areas of the humanities and the timeframe – restricted for practical reasons to two or three years – that is generally used when calculating bibliometric citation scores.

From the point of view of practice, the existing databases provide entirely insufficient coverage of publications in the humanities. There has recently been some improvement as regards this point, with monographs and non-English publications gradually being included in some databases. Nevertheless, the objections in principle continue to apply. This does not preclude that bibliometric indicators can have a certain information value in some areas of the humanities, where they can be used as proxy variables'. Nevertheless, bibliometric information – in so far as it is relevant at all – must be considered in combination with other indicators, and appraisal of all the indicators by peers is essential. There can be no question of automatic calculations.

A partial alternative to electronic bibliometric rankings is to classify output according to the status of the publication media (A, B, C journals and book series). In line with initiatives taken elsewhere (e.g. in Norway and Flanders), the Committee supports a development that it sees as an attempt to create clear distinctions in the research output of individuals and institutes. An initiative at a European level (ERIH) for classifying humanities journals was a failure and it now seems necessary to create a national classification system (with international benchmarking) for national and international media. The Committee envisages the primary role in this as being taken by the national research schools under the auspices of and directed by the Academy.

System of quality indicators

The system of quality indicators proposed by the Committee (a diagrammatic presentation is given in Section 5) is based on peer assessment (on a scale running from 1 to 5) of quality from both the scholarly and civil-society perspective and based on three criteria (output, the use made of output, and indications of recognition). A number of indicators will then be formulated for each of these criteria. Selections tailored to the particular context can then be made from these lists and they can if required be supplemented (in the 'other'category) by discipline-specific or context-specific indicators (for example mission-related indicators). There is also scope for adding new indicators that may arise due to the increasing digitisation of scholarship and publication practices. The system as a whole will consequently be flexible and as simple as possible. Selection of the relevant indicators will be left to the discipline that is being assessed (for example in the case of national external reviews) or the institute being assessed.

During the pilot phase, it will be necessary to specify further how the bodies to be assessed must provide the information relevant to each indicator, together with an investigation of how this relates to the currently normal METIS registration of research data.

Pilot phase and preliminary recommendations

The system of quality indicators drafted here will be tested during a pilot phase that will initially involve two experimental evaluations, namely at the Academy's Meertens Institute and Groningen University's Groningen Research Institute for the Study of Culture (ICOG). Amongst other things, consideration will be given to the ability of the system to make distinctions and to various aspects of practicality (for example the possibilities for 'extended peer review' and for bibliographical classification, and the most economical procedures for supplying the indicators).

During the pilot phase the Committee will also be involved with a CWTS project – commissioned by the Executive Board of Erasmus University Rotterdam – to survey the options for improved methods of bibliometric research in history in the Netherlands. The Committee adopts a neutral stance regarding the administrative and policy objectives for this project, but focuses on the question of what options the chosen new approaches present and what the balance is between extra investment and improved returns.

Conditional on the results of this pilot phase, the Committee intends making the following recommendations to the Academy in its final report:

- Take the lead in and take charge of an initiative to arrive, via the national research schools, at a broadly supported classification of the national and international publication media (journals, book series, publishers) in the humanities.
- Arrange for discussion of quality indicators to take place in the Academy's advisory councils so as to survey varying views regarding such indicators and to investigate

the possibilities regarding further consensus-forming. The latter would specifically appear to be opportune in view of the surprising convergence between the findings of this draft report and the recently published report on Quality Assessment in the Design and Construction Disciplines [*Kwaliteitsbeoordeling in de Ontwerpende en Construerende Disciplines*].

PREFACE

The interim report of our Committee on Quality Indicators for the Humanities will be introduced here with some preliminary remarks about the make-up of our Committee, its conception of the job and the structure of the report.

The following people were appointed (in their personal capacity) to membership of the Committee:

- Prof. Keimpe Algra, Chair, Professor of the History of Ancient and Medieval Philosophy, Utrecht University;
- Prof. Hans Bennis, Director Meertens Institute (KNAW), Professor of Language Variation, University of Amsterdam;
- Prof. Jan Willem van Henten, Professor of New Testament Exegesis, Early Christian-Literature, and Hellenist-Judaic Literature, University of Amsterdam;
- Prof. Mary Kemperink, Professor of Modern Dutch Literature, Groningen University;
- Prof. Wijnand Mijnhardt, Professor of Post-mediaeval History, in particular the History of Culture, Mentality and Ideas, Utrecht University;
- Prof. Paul Rutten, Professor of Digital Media Studies, Leiden University (until 31 December 2010).

The Committee was assisted by a project team consisting of:

- Dr Jack Spaapen, project coordinator, secretary to the Council for Humanities;
- Dr Koen Hilberdink, head of the Academy department for the learned society;
- Dr Mark Pen, researcher/secretary.

The Committee found itself confronted by a variety of views regarding the usefulness, need for, and possibilities of a system of quality indicators. It found, on the one hand, that the expectations of some policymakers are too positive regarding a 'simple'

and preferably fully quantified system that makes easy comparisons possible between research groups and even whole disciplines. On the other hand, the Committee noted that at the field level there is still a clear and significant aversion among researchers to the idea of 'measuring' quality and to the imposition of what are considered to be management models taken from the business sector; there is widespread concern regarding the increasing bureaucratic burden imposed by the ever-stronger culture of evaluation and accountability.[5] The Committee has attempted to avoid both these extreme positions by seeking a system of quality indicators that is as effective as possible (and therefore not merely simple) in so far as it attempts to do justice to the variety of forms and variability of scholarly practice in the humanities and elsewhere, but that is also practicable and that does not lead to an excessive bureaucratic burden. Whether the designed system actually complies with these two conditions is explicitly one of the main questions to be dealt with during the pilot phase of the project.

The Committee is aware that 'the humanities' is not a very clearly defined domain, and that its scope may differ internationally: in some countries, for example, legal studies and some areas of anthropology are considered as belonging to the humanities. The Committee adopts a pragmatic stance in this regard and follows the assignment it received from the Academy, which speaks of quality indicators 'for the humanities in the Netherlands'. It takes as its point of departure what is understood by 'the humanities' (*geesteswetenschappen)* in the Dutch context.[6]

From the start the Committee has felt that in developing a system of quality indicators it should take account of the context within which those indicators were to operate. Accordingly, the report also considers various contexts of research evaluation and peer review. The report is constructed as follows.

The first two sections offer a general introductory analysis so as to provide a framework for the system of quality indicators that will ultimately be developed. Section 1 deals with the concept of scientific/scholarly quality and the role of quality indicators in assessing research. Section 2 describes research practices in the humanities and investigates how the concept of quality should be viewed from the perspective of these practices and what this means for quality assessment and the use of quality indicators. Section 3 considers the advantages and disadvantages of peer review and the way in which quality indicators can support peer review. Section 4 focuses on the usefulness of bibliometric methods (and alternatives to them) in the humanities. The insights gained in Sections 1 to 4 create the basis for Section 5, in which a system of quality indicators is presented, together with suggestions for how that system can operate when actually assessing quality.

The Committee has based its findings on a survey of international research on quality assessments and indicators, a survey of recent developments of evaluation

5 See for example Head 2011

6 In the report of the Commissie Nationaal Plan Toekomst Geesteswetenschappen (2008: 11), the variety of fields is shown: the study of languages and culture, literature and arts, history and archaeology, religion, ethics, gender, philosophy, communication- and media studies.

procedures in a number of countries, and extensive consultation with parties in the humanities in the Netherlands. An overview of recent developments in other countries is given in Appendix 2. A list of those interviewed is given in Appendix 3.

1 QUALITY ASSESSMENT AND QUALITY INDICATORS

This section, together with the following one, offers an analytical survey of concepts, facts, and developments that are relevant to proper understanding of the problems associated with quality assessment and quality indicators in the humanities. What do we mean by scholarly quality and how is it ascribed? Why do we need quality indicators, and what are they? What do we mean by 'the humanities', and what specific features of the humanities justify their having their own specific system of quality indicators? What are the different contexts in which research evaluations take place, and to what extent do these various contexts require different approaches? The analysis offered in these first two sections on the basis of the status quo produces a number of conditions with which the quality indicators developed in Section 5 must comply. In the present Section 1, we focus on the concept of scholarly quality, on the contexts within which quality assessment takes place, and on the role of quality indicators.

The quality of scholarly research

The quality of products or services is in many respects relative or relational. It is something ascribed *by* others, it is ascribed *in relation to* certain objectives (i.e. a usage context), and it is ascribed *from a comparative perspective*. Given that quality is related to a usage context – does the product provide what the user can expect of it? – and is determined in comparison with other products that function as benchmarks, it is possible to speak of quality as something *that can be substantiated intersubjectively*.

All this also applies to the quality of scholarly research, which is also ascribed by others, in particular fellow scholars or 'peers'. It is they who determine whether the research provides what users can expect of it. The scholarly community expects new knowledge and new insights, perspectives for future research, and new discussions.

Other users – for example industry, government, cultural and civil-society institutions – in turn have their own legitimate expectations. Peers are expected to be able to articulate all – or at least most – of these expectations and to assess whether specific research complies with them. Amongst other things, they do this by utilising the performance of national and international colleagues as benchmarks. To the extent that their judgments can be substantiated, they rise above the level of the merely subjective or intuitive.

Quality assessments in various contexts

The more or less systematic assessment of the work of researchers by peers (peer review) has its roots in the eighteenth century, when researchers began communicating their results in special journals.[7] It was peer review that determined what could or could not be published. This kind of peer review, i.e. the assessment of individual publications, is still applied by the majority of scholarly journals and book series.

Nowadays, however, quality assessment of scholarly research in most contexts involves more than assessing individual publications. External reviews evaluate complete research groups and institutes. In the Netherlands, this is done according to the Standard Evaluation Protocol (SEP). Moreover such assessments are no longer merely concerned with the aggregated quality of individual publications; they nowadays include such things as the direction and management of research and the impact of research on scholarship or society.[8] The SEP also provides for an internal 'mid-term review' halfway through the six-year external review cycle, with the same criteria applying as in an official external review. Some universities are also experimenting with a type of annual monitoring by means of a 'quality card' for programmes, institutes, and faculties; this also applies a broad range of criteria taken from the SEP protocol.

Different types of quality assessment, on a smaller scale, are found when the NWO or the EU assess research proposals in the context of allocating research funding. At their request, referees assess both – prospectively – matters such as the originality of the research questions to be tackled, their relevance, the proposed method, and the feasibility and also – retrospectively – the applicant's research CV.

Finally, research groups, and researchers are assessed annually as regards their scholarly quality and productivity at virtually all faculties and institutes. This generally involves an overall assessment of the individual researcher's activities, often but not exclusively during 'performance and development' interviews in the context of personnel policy. The focus here is on assessing the research output in relation to the

7 De Solla Price (1963).

8 The SEP protocol comprises four assessment criteria: 'quality', 'productivity', 'relevance', and 'vitality and feasibility'. It should be noted that the system of quality indicators developed in the present report only broadly covers the first three of these criteria. 'Vitality and feasibility' mainly concern the strategic planning and organisation of research, in other words they do not directly affect the quality of the research as such and consequently fall outside the remit of this report. (See also the subsection on 'Application contexts' in Section 5.)

available research time. This can have consequences for the allocation or reallocation of research capacity and therefore for the duties of the individual member of staff. The criteria for such individual assessments are generally derived from the criteria in the SEP protocol. Consideration is given, as it were, to how an external review committee would view the qualitative and quantitative contribution made by the individual researcher to his or her programme or institute.

Standardisation of peer review procedures

In most of these contexts, the basis of assessment is peer review. Section 3 presents a more detailed analysis of the advantages and disadvantages of this system. For the present, it is sufficient to point out that peer review enjoys broad support as the basis for quality assessment but that the system is under pressure as regards a number of points. For one thing, the enormous growth in scholarly activity and the relative shortage of available funds have led to a significant increase in the number of evaluations and evaluation contexts (the three most important of which we have already referred to). The increasing pressure to publish is also leading to a major increase in the number of manuscripts submitted to journals and to a growth in the number of journals, whether or not in digital form. The demand for reviewers associated with all this – reviewers who are themselves under pressure to publish – is creating tension within the peer review system, with the risk of a general loss of quality. Apart from these contingent factors, peer review must always struggle against the appearance of subjectivity and the impression that it relies on intuition ('I know quality when I see it.'). Partly for these reasons, the need has arisen for standardised procedures. These can simplify the process of peer review while at the same time increasing the comparability of judgments and preventing subjectivity. To guarantee the comparability and objectivity of peer review assessments as far as possible and to streamline the procedures, initiatives have been put in place in many countries to incorporate the criteria and methods of research evaluation within a single system.[9]

Quality indicators

A system of quality indicators can substantially contribute to these systematisations and standardisations of procedures if it offers ways of substantiating quality judgments. What should one take 'quality indicators' to mean in this connection? In the context of evaluating scholarly research, the term 'quality indicators' means *all*

9 In the Netherlands, the SEP system was introduced in 2003 but the UK has had its Research Assessment Exercise (recently renamed the Research Excellence Framework) much longer. Australia has its Research Quality Framework (recently renamed Excellence Research Australia ERA), while France has the Comité National d'Evaluation de Recherche (CNER). All these systems aim to systematise quality assessment (i.e. to make it fairer and less burdensome) and effective (by making it possible to position the results within a national perspective in some way or other).

actually determinable indications of scholarly and societal quality other than the content of the scientific output itself. In that sense, the judgment of peers is itself also a quality indicator. Where the development of a system of quality indicators is concerned, however, the Committee has decided to focus on what can be referred to as *external* quality indicators, i.e. indicators outside the peer review process that provide that process with a verifiable basis**.** Some of these indicators, namely 'output indicators', reflect the standard of the products of the research concerned. Some examples are: acceptance in 'A' journals, reviews, citations, and impact that can be demonstrated in some other way. Other indicators reflect the status that a researcher or research group enjoys among his/her peers: invitations to appear as a keynote speaker at major conferences, membership of editorial boards or prestigious committees, prizes, and awards. This second category of indicators are referred to as 'esteem indicators'. A system of quality indicators will need to comprise both types.

These quality indicators, which can and must support the eventual peer review, are themselves based to a significant extent on previous peer review, and thus on qualitative judgments. Citations scores are based on decisions by colleagues in the field to cite an article; inclusion in an 'A' journal reflects a positive judgment on the part of its editors and referees; and esteem indicators also represent the judgment of others. In so far as quality indicators substantiate peer review, they therefore do so by assigning an intersubjective aspect to the judgment ("other people also think that this is high-quality research').

Recent trends

Three recent trends practice of quality assessment need to be taken into account when considering the system of quality indicators that we wish to design: the advance of bibliometrics, growing attention to societal relevance, and the diversification of assessment contexts.

In order to provide not only intersubjective but also quantitative substantiation for the quality judgments of peer review, increasing use has been made in recent decades of the system of bibliometrics that was developed in the 1970s. Originally developed to clarify processes in the context of the sociology of science, bibliometric tools have gradually taken on a different role. They are now increasingly used as indicators intended to make clear, on a quantitative scale, the publication output of scholars and the relative quality of that output.[10] Examples include impact factors assigned to journals, citation statistics, and various citation indexes. The bibliometrics experts themselves are generally more cautious in their claims regarding the usefulness and

10 The database of the Institute for Scientific Information (ISI) on which the first bibliometric studies were based comprised only publications in the natural sciences. Initially, bibliometrics took no account of the humanities. That has been changing somewhat in recent years, but it is a slow process. Section 4 deals in greater detail with the relationship between bibliometrics and the humanities.

significance of these tools than some policymakers in government, at universities, and within national and international bodies that finance research. Policymakers tend to have a simplistic idea of what bibliometrics involves, seeing it as a simple means of generating objective rankings of researchers and research groups by just pressing a key. Conversely, some researchers tend to unthinkingly reject everything that has to do with measuring 'quality.' Accordingly, a balanced judgment is called for. Section 4 will therefore focus quite extensively on the usefulness of bibliometric methods in general and in the humanities in particular.

Another recent development, since the 1990s, has been the trend towards assigning increasing importance to the societal relevance of science and scholarship. This has come to play a role nationally and internationally in many evaluations of quality, and it is now a standard component of the framework offered by the SEP for evaluations in the Netherlands. This trend can also be observed in the humanities. There is an increasing realisation within both institutions and government that the humanities make an important contribution to what the French sociologist Pierre Bourdieu has referred to as 'cultural capital' (knowledge, skills, education).[11] One can also consider the contribution made by the humanities to public debates, for example on history and identity, and to policy-making in the area of culture and the media. From a methodological point of view, assessing the societal quality of research is more complicated than assessing scientific/scholarly quality. The data are less robust and sometimes more complicated to collect, and it is not a matter of course that the relevant information can always be estimated at its true worth by one's colleagues in the same field;[12] moreover, many different contexts are involved. Nevertheless, the view that quality is determined in part by the usage context obliges us to take account of societal quality, and to do this on the basis of separate indicators.[13]

11 See AWT (2007) 5–6: 'Knowledge in the humanities and social sciences is in great demand in such fields as the law, education, mental health care, politics, and policy. The general public also show great interest, as appears, for example, from sales of management books and historical studies and from the number of visitors to museums or heritage sites. Research in the humanities and social sciences is also essential when it comes to policy development. Issues regarding cohesion and integration, inter-generational solidarity, international relations, market processes, globalisation, and education reform could not be tackled without the aid of the humanities and social sciences. Even technical problems, for example concerning mobility, safety or climate change, have a great need for the contribution made by these disciplines. These problems cannot, after all, be solved without people changing their behaviour.'

12 In assessing the societal quality of research, it may therefore be useful to make use of 'extended peer review' (see Section 3).

13 By emphatically giving societal quality a place in its reporting, the Committee links up with the report *Impact Assessment Humanities* [*Impact Assessment Geesteswetenschappen*] published by the Netherlands Organisation for Scientific Research (NWO) in 2009 and the *Guide for Evaluating the Societal relevance of Scientific Research* [*Handreiking Evaluatie van Maatschappelijke Relevantie van Wetenschappelijk Onderzoek*] resulting from the ERiC project involving the Academy, the NWO, and the Association of Universities in the Netherlands (VSNU) (2010).

A third relevant trend has already been referred to in the present section, namely the large number of research evaluations that take place nowadays, in various different contexts and with various different objectives. Universities and organisations such as the NWO require data that they can use to arrive at responsible decisions on prioritising and allocating funds. Researchers and their assessors wish to know how they are doing compared to the competition (including internationally). Local administrators are concerned about whether their institute is fulfilling its aims and whether some groups are achieving better results than others. A practicable system of quality indicators will need to be able to be used in these various different contexts. Not all of the indicators are relevant in every context, or at least not equally relevant. We therefore require a fully equipped 'toolbox' for making selections for various different purposes and contexts.[14]

A fourth recent trend also needs to be considered here. Digitisation of the primary research process and the way in which research knowledge is shared between peers and with society in general can potentially bring about major changes in the system of research production, with all the possible consequences that may have for ideas regarding quality and quality indicators.[15] Digitisation makes processes more transparent, including processes of scientific and scholarly research. More than in the past, quality as a label can also be attached to these processes themselves and not just to end products.[16] Digitisation also generates ways of sharing knowledge during the research process, for example by means of scholarship/science blogs that earn recognition as components of research output and that can indicate quality. All this means that an exclusive focus on finalised articles in peer-reviewed journals, and in general on a fixed text as the acme of research production, is no longer necessary.[17] As yet, the implications of digitisation for research practice are still too vague to be converted into a specific system of quality indicators, but the system to be developed must offer scope for additions and adjustments on the basis of these new developments.

14 By emphasising the importance of a flexible system with the option of including context-specific indicators – related to the specific character of a given discipline or the mission of a given institute – the Committee endorses the findings of the Academy's report *Judging Research on its Merits* (2005).

15 See, for example, Borgman (2007).

16 Semi-finished products, for example a provisional database, can therefore also be the object of quality assessment. See Verhaar et al. (2010).

17 The practice of 'open review', for example, assumes that research results are first shared digitally and reviewed openly. Authors then amend their contribution on the basis of the criticism received. Further research by the original authors can then lead to new versions, and in the course of development an external commentator can join them as a co-author. All this means that the production process becomes more open, dynamic, and communal.

The first three of the trends that we have noted show that a tension exists between simplicity and effectiveness.[18] Bibliometrics has gradually been refined and has thus become more complicated. Taking account of societal relevance is important but it is no simple matter. A system of quality indicators that can be used in a variety of different evaluative contexts must therefore be not just wide-ranging and fine-meshed but also flexible. In short, an effective system of quality indicators cannot be a simple one, and a simple system will give only crude indications. One should therefore aim for an optimum balance

Conclusions

The quality of the products of scientific/scholarly research can basically be determined by the researcher's peers. Quality assessments by peers can be carried out in a variety of different contexts and with various different intentions, both prospectively and retrospectively. Retrospective quality assessments can be substantiated by making use of quality indicators. These support the process by giving peer review a broader and intersubjective basis. At the same time, they make the process more transparent and thus in a certain way simpler: not by introducing simple calculations based on one or more indicators but by introducing structure and comparability.

CONDITIONS

We will deal at length with the system of peer review and the usefulness of bibliometrics later on (in Sections 3 and 4). For the rest, the findings of this first section allow us to formulate the following conditions for the system of quality indicators that is to be developed:

- A system of quality indicators must comprise both output indicators and esteem indicators (indicators that express previous assessment by peers and recognition by the scientific community);
- A system of quality indicators must be wide-ranging enough to take account not only of scientific/scholarly quality but also of societal relevance;
- The scope of a good system of quality indicators should not impair practicality;
- The system must be flexible enough to make a tailor-made approach possible in specific contexts;
- The system of quality indicators must be open to adjustment or revision once the consequences of digitisation for research practices have become fully clear.

18 The Academy's instructions to the Committee refer in this connection simply to the development of 'a system of quality indicators'. The *Sustainable Humanities* report on which those instructions are based speaks, however, of the 'simplest, clearest, and most effective system'. Given this tension between simplicity and adequacy, the Committee has interpreted this as a request to produce the optimum balance between simplicity and precision.

2 RESEARCH AND RESEARCH QUALITY IN THE HUMANITIES

This section describes a number of aspects of research practices in the humanities. It begins by considering the extent to which the humanities, as a field, display a sufficiently specific character of their own as to require an area-specific system of quality indicators. Reasons will then be given as to why the internal diversity of the humanities demands a wide-ranging, flexible system of indicators.

The specific character of the humanities

The humanities focus on human culture and its past and present products: language, institutions, religion, philosophy, literature, the visual arts, architecture, music, film, and media. In the past few decades, the contribution from an ideological perspective (gender, race, class) and an ethical perspective and more institutional and sociological approaches have led to increasing interdisciplinarity, also vis-à-vis disciplines outside the humanities. The humanities have to to a considerable extent developed independently of one another. They display a kind of 'family resemblance' but they do not share a common 'essence'. This means that it is impossible to specify general characteristics of 'the' humanities. Different countries do not always classify the same disciplines as belonging to the humanities. In the Dutch context too, such terms as *geestesweten-schappen*, 'humanities', or 'humaniora' do not always have the same scope, and the institutional categorisation of such subjects as theology, philosophy, and archaeology may differ.

There is no single common feature distinguishing the humanities from other fields of research. Rather, the humanities are a conglomeration of various different disciplines, each with its own place within the totality and with variable cross-connections

with other fields. The methods of some disciplines within the humanities, for example linguistics, are very similar to those in the empirical natural sciences. Theories are created and modified, hypotheses are tested empirically, and there is cumulative production of a shared 'body of knowledge'. Other disciplines within the humanities, for example some areas of history, are more closely related to certain branches of the social sciences. Yet others focus mainly on interpreting individual or unique objects (works of art, literary or religious texts). Some branches of philosophy display similarities with areas of mathematics.

The fact that the humanities do not constitute a single entity with a shared identity means that one cannot adduce such a presumed identity as a reason why quality assessment of the humanities as a whole – i.e. as the humanities – should be any different to quality assessment in other fields. Actual research practices support this conclusion. Despite all the differences in methodology and research practices, there is far-reaching agreement within the humanities regarding what scholarly research aims to achieve. that does not differ significantly from the aims of other areas of research. As in other fields, scholarly research in the humanities involves the methodical increase and clarification of our knowledge of reality and of our rapport to that reality. And as in other fields, the aim of research can also comprise societal relevance: scholarly knowledge is often shared with a broader target group made up of policymakers, cultural institutions, educators, and other interested parties. In all of these different humanities, research quality in the general sense is viewed in the same way as in other fields. In the humanities too, quality applies in the sense of the significance of the results of research for the scholarly community and for others. Here too, it is primarily the researchers' peers who are considered able, partly with the aid of indicators, to judge the quality of the research. A large number of the quality indicators that apply in other fields can therefore automatically be utilised for the humanities too. All the various disciplines within the humanities have leading journals, important conferences to which researchers can be invited as keynote speakers, awards, and prizes. Seen from this perspective, the system of quality indicators that is to be designed will not be significantly different to what one could create for other scientific/scholarly fields.

Diversity within the humanities

The humanities therefore do not require their own *type* of quality indicators. However, the internal diversity of the humanities – a diversity that can also be found in other areas of research – does demand a wide *range* of indicators. Three kinds of diversity – differences as regards the sociology of science, differences in objectives and products, and differences in publication channels – need to have their effect in the form of a broad spectrum of quality indicators, from which one can choose according to the particular discipline and sometimes according to the particular context.

Differences as regards the sociology of science can lead to certain indicators being more relevant in one discipline than in another. In one field, for example, a large

number of researchers may be working simultaneously to develop and refine the same theory, while in another individual researchers will be working on various different topics. This makes a major difference to the applicability of bibliometric indicators (for example citation analyses) and it sets limits to the bibliometric comparability of various disciplines (and component parts of disciplines).

There may also be relevant differences between the objectives and products of the research. Those working in a number of humanities disciplines are the guardians, access-providers, and interpreters of national and international heritage. As with research in some other fields (for example the technical sciences), humanities research does not 'translate' exclusively into the 'classic' form of articles in scientific periodicals but also into other products. Many researchers are engaged in constructing databases and providing access to collections of data, sometimes for fellow researchers and sometimes for a broader public. In doing so, they make an important contribution to building up a knowledge infrastructure in the humanities. This means that for quality assessment in some disciplines in the humanities there is a need for a wider selection of relevant indicators than scholarly articles alone.

A third point concerns the publication channels that are used. In the case of the more synthesising humanities, journals, monographs, and collections dealing with particular themes are an important channel of communication. In many disciplines within the humanities, the list of a leading researcher's publications includes at least a few high-profile scholarly books. For most humanities disciplines, monographs are therefore relevant indicators.[19] But in those cases where monographs are relevant quality indicators, bibliometric indicators are often in fact of less relevance because they are as yet based mainly on databases which do not include monographs. In addition, some branches of the humanities concern themselves with an object of study that is geographically so specific – for example Dutch regional history, Italian or Hungarian poetry – that English is by no means the obvious language for publishing research. In order to count internationally in such fields, it is necessary to publish in languages other than English. In these cases too, bibliometric indicators are of only limited use, at least to the extent that they are still based exclusively or primarily on citations of articles in English-language journals.[20]

Conclusions

In their striving for both scholarly quality and societal relevance, the humanities are comparable with most other fields of science and scholarship. This also applies to the way in which the concept of 'quality' is applied in actual practice. The humanities therefore do not require their own *type* of quality indicators.

19 These may also include digital monographs, perhaps increasingly. See Adema (2010).

20 See Section 4 for further discussion of the limits to the usefulness of bibliometric indicators.

An effective range of criteria and indicators will, however, need to be broad enough to do justice to the diversity of approaches, objectives, target groups, and publication channels that are to be found within the humanities. A system of indicators that confines itself almost entirely to measuring research output in the form of English-language articles in journals published in the English-speaking world that can be fruitfully investigated using bibliometry is inadequate where most disciplines within the humanities are concerned.

The Committee wishes to emphasise once more that many other areas of science and scholarship display a similar diversity, and that a limited system of indicators is inadequate in those areas also.

CONDITION

The findings of this second section produce the following condition as regards the system of quality indicators that is to be designed:

- The humanities require a wide range of quality indicators with a view to the diversity in
  - research practices and publishing cultures;
  - publication languages and publication media;
  - research products;
  - target groups.

The diversity within the first two of these imposes limits on the applicability of bibliometric indicators. The final two points are complementary and require a relatively broad range of quality indicators.

3 PEER REVIEW AND QUALITY INDICATORS

It was asserted above that quality indicators can assist and supplement peer review. It is now relevant to investigate in which respects peer review requires such substantiation and what limits are set to the applicability of quality indicators in this connection. We have also seen that quality indicators are based on forms of peer review. This means that relevant criticism of peer review can also call the value of quality indicators into question. These are two reasons to include a brief consideration of peer review. The present section will discuss how peer review is applied, the aspects that are open to criticism, how that criticism can be met, and the consequences of all this for the role and value of quality indicators.

Peer review: acceptance and limitations

There is great support within the scientific community for peer review as a method for assessing quality and peer review is utilised in all major evaluation contexts. Ideally, the judgment in peer review is given by experts and it is intersubjective and argued. However, the various ways in which peer review can be organised involve that these advantages are not always present to the same extent.

The peers involved are sometimes indeed experts, but in some cases broad panels give the judgment, whether or not supported by referee reports from specialised colleagues. The procedure for selecting peers is not always transparent and the peers are not always what they in fact ought to be, namely experienced top-class researchers who themselves have a high score as regards the indicators that they are required to apply and who are able to take a keen look at sample publications and research proposals and to assess indicators for what they are worth and balance them against

one another.[21] Peer review is sometimes indeed collective – involving an external review committee, an assessment panel, or several referees – but in other contexts it is only a single individual who gives his or her assessment: in some cases an article may be assessed by a single referee and the progress of a researcher's work is often assessed by only a single research director. The reasons supporting a peer review judgment are also not always equally extensive and explicit, and in some cases no reasons are in fact given; this detracts from the value of peer review.

A certain leeway as regards the points of intersubjectivity and discipline-specific expertise may be acceptable in practice, depending on the context and purpose of the assessment and the possibility of procedural compensation.[22] Within that acceptable leeway, however, various things may go wrong. Different peers can arrive at different conclusions on the basis of the same material because of differences in their background, knowledge, preferences, or effort ("sloppy refereeing'). In the literature and in recent discussions, one accordingly finds a number of critical remarks regarding the usefulness and reliability of peer review. The most important of these are listed below:[23]

1. Peer review may be based on *subjective preferences* for a certain type of research, particularly when the criteria applied are frequently implicit.
2. Peers may have a tendency to give preference to work that *links up with existing views and paradigms*, something that may act as a brake on controversial research and innovation.
3. Peer review can have negative consequences for *interdisciplinary research* taking place at the boundaries between different disciplines.
4. In a small scientific/scholarly community, *shared interests and mutual dependence* may lead to non-intervention. The peer who today assesses a researcher's article may be himself assessed tomorrow by that same researcher. The peer's judgment may then be a 'false positive'.
5. There may also be *conflicting interests*. The peer may hinder or delay research that is very similar to his own research so that he can be the first to publish his results. In that case, the peer's judgment may be a 'false negative'.
6. Peers may have a preference for the work of *established researchers*, leading to an accumulation of positive judgments and researcher funding for researchers with a good reputation. This is referred to as the 'Matthew effect' ("For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be

21 This is undoubtedly connected with the increasing need for peer review referred to in Section 1.

22 Individual and consequently non-intersubjective peer review by a research director or manager may be acceptable because it is normally restricted to general points – is the person concerned functioning as he/she ought to? – and can otherwise rely on judgments by external review committees or mid-term panels or other information. Broader, non-discipline-specific panels can be acceptable when supported by reference reports that are in fact written on the basis of discipline-specific expertise (NWO).

23 See for example Hemlin (2009).

taken away even that which he hath.' Matt. 25:29).[24]

7. In many cases, the procedure for selecting peers is not transparent and may involve manipulative aspects.
8. As *fellow scholars*, peers are not by necessity capable of assessing the *societal* quality and the impact of research.
9. One final objection is of a more practical nature, but still important. The increasing stress on competition and accountability in the research sector involve that leading peers must devote more and more time to assessing researchers and research proposals. This leads to excessive pressure on certain peers, with the risk that they will not always assess articles and proposals seriously. This can create problems as regards the reliability of the peer review process.[25]

Consideration of the criticisms

Not all of the above criticisms of peer review are relevant in all contexts, nor are they all of the same importance. Criticisms (4) and (5) involve the peer review of individual publications and have to do with integrity. Lapses in the assessment of individual publications do naturally occur, but the Committee has the impression that this type of peer review normally takes place in good faith. Moreover, peer review of individual publications is generally expressed in a reader's report, thus enforcing objectivity. In many cases, there are also several referees. The risk would therefore seem acceptable, and there is no better alternative. Criticism (6) concerns peer review of research proposals and grant applications. It is not clear, however, how much damage the 'Matthew effect' causes in such a context or whether it is indeed a serious problem. What it basically comes down to is, after all, that a researcher is given credit for his/her performance in the past, which – to a certain extent – is precisely the intention of this type of research assessment.

The general criticisms (1) to (3) would appear to be more serious. There are, however, procedural methods available to meet these objections. Criticism (1) can to a large extent be neutralised in the context of the assessment of individual publications by requesting a reader's report; in contexts in which an entire CV or a whole group is being assessed, one can request that the assessment be substantiated by intersubjective quality indicators.

The risks involved in (2) and (3) can be reduced to an acceptable minimum by making alterations to the procedure that is followed. Certainly in cases involving broader panels or several referees, selection of the peers can be used to create openness to innovation in the relevant research domain and openness to interdisciplinarity. Peers with an excellent reputation in a highly specialised discipline can be supplemented by peers with a broader and more interdisciplinary perspective. One can also deal with criticism (7) by making panels broader and by avoiding review by only a single peer.

Criticism (8) can be dealt with by working with 'extended peer review', i.e. peer

24 Merton (1968) 56–63.

25 See for example the Academy's Quality Assurance Committee (2008).

review supplemented by information from a wider range of experts than just colleagues within the same discipline as the researcher. Societal stakeholders can be brought into the assessment process, for example, either by means of external reporting or by assigning them a role on the assessment panels in addition to the actual peers.[26]

Where criticism (9) is concerned, the Committee wishes to note that it is essential to prevent excessive pressure on peers. A reduction in the number of assessments and simplification of the relevant processes are highly desirable. One can attempt to prevent duplication and to make use of the results of evaluation in more than one context (as happens, in the Dutch context, with the results of the SEP evaluations when research schools are accredited by ECOS). A clear system of quality indicators enjoying broad support can also simplify the work both of peer review committees and the researchers and research groups that are being assessed, viz. by providing structure. A condition for this, however, is that the system is flexible enough for a tailor-made approach to be adopted in specific contexts; in addition, there should be enough of a fit between the relevant quality indicators and the systems for recording research output (for example METIS). Section 5 will deal with these conditions in more detail, as well as with the options for creating them.

Peer review as the basis for quality indicators

Are the criticisms levelled at peer review so serious as to affect the value and reliability of the various types of quality indicators that are themselves based on peer review (for example acceptance by 'A' journals, citations, reviews, various esteem indicators)? The Committee believes that they are not. In the first place, the indicators referred to are generally based on *aggregation* of individual peer review decisions, meaning that subjective bias is neutralised as far as possible. Secondly, in actual practice indicators do not operate individually, but in combination with other indicators, consequently guaranteeing a sufficiently broad basis for assessment. Although themselves ultimately based on peer review, quality indicators can therefore be used in their turn to support and substantiate peer review processes. As we have seen, they do this by broadening the intersubjective basis for assessment.

Conclusions

Most researchers accept peer review as the core of the processes of research quality assessment and the Committee too considers that peer review, imperfect as it may be, is the best method that we have available. In applying it, we must, however, allow for certain risks; these can, in part, be reduced by means of guarantees within the

26 See ERiC (2010). The possibility and practicality of bringing societal stakeholders into the process of peer review is one of the aspects that will be tested during the pilot phase of this project. (See also Section 5.)

procedures followed. In addition, a system of intersubjective quality indicators can be used to support the process. These reflect, as it were, the accumulation of individual peer review judgments and can consequently reinforce and substantiate the assessment provided by a single individual or a committee. Provided that it complies with certain conditions regarding practicality, a clear, broadly applicable system of quality indicators can also simplify the work of both peer review committees and the groups being assessed.

CONDITION

A new condition is thus produced with which a system of quality indicators must comply:

- A practicable system of quality indicators should be applicable with maximum ease, both from the point of view of the institutes that are required to provide the relevant information and from that of the peers whose task it is to weigh up the various indicators.

4 BIBLIOMETRIC METHODS

Bibliometric databases were first compiled in the 1960s. Originally, they were intended to be able to trace information more rapidly within the growing volume of literature and for the purposes of research on communication patterns and collaboration arrangements in science and scholarship. The idea was that examining patterns of citations would enable one to understand trends in the sociology of science. In the course of time, however, the databases also came to be used to assess the quality and impact of research. Currently, citations in the biomedical and natural sciences are taken to be a more or less direct indication of research quality, with frequent citation now being considered synonymous with scientific quality. Other fields of research are under a certain pressure to also subject themselves to assessment using bibliometric indicators. As we have already seen, the advantage of bibliometric indicators as seen by many policymakers is their objectivity and supposed lack of ambiguity. The bibliometrics experts themselves generally adopt a much more balanced view regarding the direct and simple application of such indicators in assessing research. For both theoretical and practical reasons, one can question their large-scale, ill-considered use. This section discusses a number of bibliometric tools and sets out what bibliometry can and cannot currently say about research in general and about quality in the humanities in particular.

Bibliometrics, impact factors, Hirsch index

Bibliometrics experts count publications and citations and then make statements in order to determine productivity and the extent to which researchers are cited. They base their counts on databases containing articles from a large number of scholarly

journals, such as Thomson Reuters' Web of Science (WoS).[27] Various types of calculations generate different quantitative indicators, for example the number of citations per publication or relative citation scores in the discipline within which publication takes place. Citation scores can naturally vary depending on the timeframe applied (i.e. the number of years over which measurement takes place).

Another commonly used quality indicator is the impact factor of the journal in which an article is published. The impact factor shows how often articles in that journal have been cited (often calculated over the preceding two years). An impact factor of 3 means that the articles in the journal concerned were cited an average of three times in years t-1 and t-2. An impact factor represents an average and therefore says nothing about the range between articles cited frequently or infrequently. It is also the case that journals that appear weekly and that cover topics that are scientifically 'hot' (for example *Nature* and *Science*) have a significantly higher impact factor than those that appear much less frequently, and for relatively small specialised disciplines.[28]

Motivated by his dissatisfaction with the existing methods of measuring citations, the physicist Jorge E. Hirsch introduced a relatively simple new method in 2005, the 'Hirsch index' (or 'h-index'). This is a discipline-dependent index that shows the 'career impact' of a researcher. Citation scores for a researcher are often distorted because a small number of the researcher's articles are cited extremely frequently. The *h*-index – which is based on the same databases (for example Web of Science) as the standard citation statistics – works from the impact of the researcher's whole oeuvre. A researcher has index H if H of his/her total of N publications have been cited at least H times in other publications and the other publications (N-H) have been cited fewer than H times. For example: an *h*-index of 10 shows that 10 of a researcher's publications have been cited at least 10 times. Hirsch argues that this index gives a more realistic picture of the impact of somebody's publications than other statistics such as the total number of citations – which can be influenced by a small number of much-cited articles, for example survey articles or reviews – or the number of citations per article. The downside of Hirsch's system is that a highly influential researcher who only publishes relatively little but who is cited frequently will have a low *h*-index. The *h*-index is derived, after all, from both impact and productivity. It also disadvantages young researchers and it is not consistent: rankings do not remain the same if two researchers who are being compared receive extra citations to the same extent (a problem that also occurs with the Journal Impact Factor).[29]

The *h*-index has become perhaps the best-known and most frequently utilised bibliometric indicator. Various other multidimensional indexes have now been developed,

27 Web of Science used to be the property of the Institute for Scientific Information (ISI) and was known by the names 'Science Citation Index', 'Social Science Citation Index', and 'Arts and Humanities Index'.

28 Journals such as *Nature* and *Science* have an impact factor of about 30, whereas the most important journal in physics (*Physical Review Letters*), for example, has an impact factor of about 7.

29 Waltman & Van Eck (2009)

either to complement the *h*-index or to allow more precise measurements for other specified contexts. We now have G, A, R, and AR indexes, for example.[30] The system of bibliometric indicators is thus subject to constant expansion and refinement.

The value of bibliometric indicators

Caution is required in a number of respects when utilising bibliometric indicators to assess research quality. Simple unweighted counting of *publications* in fact provides only a measure of productivity and not of quality. Counting *citations* is already more informative, but it also fails to provide any direct indication of quality; at most, it gives an indication of impact. Even if one corrects for self-citations (and also leaves aside possible 'citation clusters' made up of researchers who are on friendly terms with one another), it is still the case that referring to someone else's work only indicates that the researcher concerned has made use of that work. Whether the reference has been positive or negative is not apparent. It can of course be argued that even a critical reference still indicates that the work cited was in any case considered important enough to be challenged and there is certainly something to be said for taking account of this type of impact in quality assessments. There is, nevertheless, no one-to-one correspondence between a high citation score and high quality.

This becomes even clearer when one considers that there are significant differences in citation patterns and citation cultures between different disciplines and even between parts of one and the same discipline. In some fields, numerous researchers carry out work on the same types of research questions, which logically leads to heavy 'citation traffic'. Specialist subdisciplines or topics studied by only a few scientists or scholars generate far fewer citations. In some disciplines it is more 'the done thing' to give numerous citations than in others, and there would also appear to be a difference between the practices in the English-speaking world and the French-speaking world, for example.[31] Such differences make it a risky matter to carry out comparisons in an evaluative context between the citation scores of researchers, and the impact factors of journals, from different research fields. One also needs to take account of the differing scores of different *types* of publications. Review articles, for example, have a greater citation impact than ordinary articles. Survey articles and handbooks are also probably cited more frequently than specialist studies. In all these cases, a high impact is not equivalent to high quality.

In the light of these considerations, we therefore need to be sceptical regarding the automatic, ill-considered use of citation scores as a *direct* and unambiguous indication of the quality level of researchers or research groups.

An additional problem in applying bibliometric methods in the humanities is that the 'half-life' of publications in many of the humanities is significantly longer than in the natural sciences, for example: the impact of a publication continues for a much

30 See, for example, Jin et al. (2007) 855–863; Costas & Bordons (2008) 267–288.
31 Dolan (2007) 28.

longer period and displays far less of a peak in the first few years after the publication has appeared. This means that bibliometric counts using the timeframe of only a few years that is generally applied in the natural sciences can give only an inadequate idea of the actual impact of publications in the humanities.

In addition to these more theoretical caveats, there is a more practical but highly important limitation. Bibliometric indicators provide their insight – a limited insight, as we have seen, and one that certainly requires interpretation – only if the coverage of the underlying database of publications and citations for a researcher is sufficient. That is by no means always the case and it certainly does not apply to most areas of the humanities. The databases of Thomson Reuters' Web of Science, which are the basis for most bibliometric analyses, do not completely cover the literature of the humanities, primarily because they include neither books nor non-English journals. The table below gives an overview of references to articles in Web of Science journals for various different disciplines. In the humanities (at the bottom of the table) only one third of the references concern articles in journals (the others concern books or collections) and only half of those journals are included in the Web of Science. More than eighty percent of the references in the humanities literature are therefore to publications (articles or books) that are not included in the Web of Science. As the table shows, the Web of Science provides reasonably good coverage mainly for the biomedical and natural sciences.

Table 1 References to articles in Web of Science journals 2002[32]

Discipline	References to publications in journals	Proportion of WoS journals in the references to publications in journals	Significance of WoS (column 1 × column 2)
Molecular biology and biochemistry	96%	97%	92%
Biomedical sciences (human)	95%	95%	90%
Chemistry	90%	93%	84%
Clinical medicine	93%	90%	84%
Theoretical physics and astronomy	89%	94%	83%
Applied physics and chemistry	83%	89%	73%
Biological sciences (plants and animals)	81%	84%	69%
Psychology and psychiatry	75%	88%	66%
Geology	77%	81%	62%
Other social sciences/health sciences	75%	80%	60%
Mathematics	71%	74%	53%
Economics	59%	80%	47%
Technical sciences	60%	77%	46%
Other social sciences	41%	72%	29%
Humanities	34%	50%	17%

32 Moed (2005) 126.

In the humanities disciplines, the coverage percentage ranges between 20% (archaeology) and 7% (Dutch studies and German studies).

A recent study by the CWTS at the humanities faculties of the two Amsterdam universities shows a similar picture. Here too, Web of Science misses out the greater part of scholarly production in the humanities. In contrast to the high coverage percentages (more than 85%) for the natural and biomedical sciences, the humanities achieve no more than 27% and 22% at the two universities respectively.[33] This means that citation counts based on the Web of Science currently give an unreliable picture of the humanities.

Efforts are being made at various levels to do something about the inadequacies of the databases. In the past six years, Thomson Reuters has increased the number of humanities journals covered by some 30%.[34] Since 2009, Web of Science also includes 1228 'regional' journals, defined as 'journals that typically target a regional rather than international audience by approaching subjects from a local perspective or focusing on particular topics of regional interest'. Half of these journals are of European origin.

An alternative to Web of Science is Elsevier's Scopus database. The level of coverage is higher than that of Web of Science but the precise extent of the coverage is unclear and researchers' addresses are not given, meaning that the database is not suitable for consistent analyses. Competition between Web of Science and Scopus will probably lead eventually to better coverage of journals in the humanities.[35] At the Frankfurt Book Fair 2010, Thomson has announced a citation index for books, but it is doubtful whether this project will be implemented in the foreseeable future.

A third database is that of Google Scholar, which also includes books and sections of books. Google Scholar is not, however, transparent as regards the content of the database. Bibliometrics experts and researchers do not know of which journals are searched by the search engine or which books are included in the databases. This makes it impossible to know how up-to-date and extensive Google Scholar search results in fact are.[36] Google Scholar is therefore also unsuitable (at present) as a bibliometric database for the humanities.

All in all, the conclusions that we can draw regarding the use of bibliometric indicators in the humanities are not unequivocal. There are, on the one hand, no reasons of principle why bibliometrics should never work anywhere for the humanities. In some fields, bibliometric indicators can in fact provide useful information. There are some fields, however, in which the publishing culture is such that bibliometrics is of little use. In addition, as we have seen, there is the problem of the inadequate coverage of the databases used, although that situation looks likely to gradually improve. Finally,

33 *idem*, 130.
34 Moed et al. (2009).
35 Hicks & Wang (2009).
36 Ibid. See also Harzing (2008).

there is the basic point that bibliometric scores cannot be converted directly into judgments of quality. A great deal of further research is still required on the validity and usefulness of the various bibliometric tools.

All this means that bibliometric indicators are not equally useful in all fields and that they must certainly not be used in isolation as a standard indicator. Where they are applicable at all, they must always be used in combination with other indicators and they must be appraised and interpreted by peers.

Classifications of journals and scientific publishers

The problems associated with applying the more common types of bibliometrics to research assessments in the humanities have led to the use in various places of bibliographical lists as an alternative system for classifying the output of scholars in the humanities. Such bibliographical classifications can play a bibliometric role in so far as they offer evaluation panels the possibility of attaching weighting factors to researchers output.

The European Science Foundation (ESF) took the initiative a few years ago to set up the European Reference Index for the Humanities (ERIH), which was intended 'to be both a bibliographic and a bibliometric tool, i.e. to facilitate both access to and assessment of humanities research. A further key aim was to raise the threshold standards of editorial practices of journals throughout Europe'.[37] This formulation suggests that the ERIH was also originally intended to act as the basis for a digitised bibliometric tool, probably in the form of a database that for the humanities would offer an alternative to the Web of Science.

The ESF invited panels of expert in fifteen disciplines to draw up lists of journals that are important as publication channels for the humanities in Europe. In order to be placed on the ERIH list, a journal must qualify as a 'good scientific journal' and pursue an objectively determinable peer review policy. This means that (as with inclusion in the ISI reference index) being placed on the ERIH list is in itself a quality indicator. Each journal is also allocated A, B, or C status within a discipline.[38] The fifteen lists currently contain a total of some 5200 journals.[39]

Whether the ERIH can grow to become a European bibliometric counterpart to the Web of Science indexes is open to doubt. It is even questionable whether the list will ever operate as a generally accepted bibliographical tool and whether the

37 European Reference Index for the Humanities (online).

38 The categories are defined as follows: A: 'high ranking international publications with a very strong reputation among researchers of the field in different countries, regularly cited all over the world'; B: 'standard international publications with a good reputation among researchers of the field in different countries'; C: 'research journals with an important local/regional significance in Europe, occasionally cited outside the publishing country though their main target group is the domestic community.'

39 Of these, 360 are relevant within two disciplines, 33 in 3 disciplines, and 5 in 4 disciplines. The list also includes Dutch language journals.

classifications applied will be adopted as quality indicators. A great deal of criticism has been expressed regarding the way in which the list was drawn up. It would appear from various sources – including interviews with the Dutch researchers involved in drawing up the ERIH list – that the inclusion of certain journals was too arbitrary. Initially, there was a one-sided focus on journals published in the English-speaking world, but the tide later turned and discussion came to be dominated by the more political question of the representation of large and small language areas. According to many people, too many national journals of doubtful or at least less than top quality were included at too high a level on the scale on the recommendation of committee members from the relevant countries. Checking of the rankings assigned by individual committee members also left something to be desired. A great deal of weight was in fact given to the judgment of only a single committee member for each research field. Partly as a result of this, the categorisation into A, B, and C journals has since remained controversial.

National systems of classification

Most disciplines have since set aside the ERIH list. A quick survey showed that in the Netherlands it is hardly used at all. The above already shows that one major reason for the failure of the ERIH initiative was that an attempt was made to produce a single European list on which all the different language areas and journals would be represented. It was perhaps a case of wanting to do too much all at once. For that reason, and also because of the lack of an overall bibliometric or bibliographical database, Hicks & Wang[40] (2009) suggest in a study commissioned by a consortium of European bodies that fund research that national systems should be developed.

Independently of that recommendation, national classifications for publication media have in fact been developed in countries including Belgium (Flanders), Denmark, and Norway. In Norway, the Norwegian Association of Higher Education Institutions set up a national research database in 2004 containing the bibliographical details of all significant academic publications in all fields of science and scholarship, i.e. including the humanities.[41] The database is used in order to distribute national research funding to scientific and scholarly institutions. The types of publication that it recognises are articles in journals, articles in collections, and monographs. The publication channels are classified as belonging to the basic level ("level 1') or a select level ("level 2').[42] This leads to points being allocated at two levels: a monograph receives 5 points at level 1 and 8 points at level 2; for an article in a journal, 1 or 3 points are allocated

40 Hicks & Wang (2009).

41 Norwegian Association of Higher Education Institutions (2004). 'Local' publications – defined as publications for which more than two thirds of the contributions come from the same institute – are not included in the system.

42 Criteria for 'level 2': (a) being perceived as the leading publication channels in a wide variety of academic contexts; and (b) publishing the most outstanding works by researchers from different countries.

respectively, and for an article in a collection 0.7 or 1 points. Because the criteria for allocating the label 'level 2' vary from one field of research to another, the system comprises three types of disciplines ('academic field groups'), each with its own set of criteria. The classification system was introduced in 2005 and utilised for the first time in the allocation model in 2006. It involves weighted counting of publications and makes no use of citation scores.[43]

For similar reasons – i.e. the distribution of research funding to institutions – work began in Flanders in 2009 to construct the Flemish Academic Bibliographic Database for Social Sciences and Humanities (VABB-SHW).[44] In the new (2008) funding system, this involves the 'variable research component' of 28% of the funds to be allocated to the Flemish universities. Of this, 30% is determined by the research output. Initially, the intention was to distribute this portion according to the allocation formula applied by the Special Research Fund (BOF), which has in fact been based since 2003 on the output (and citations) of mainly the biomedical and natural sciences because a decision had been taken to base matters on the databases of the Web of Science. As already mentioned, those databases cover only a relatively small proportion of publications in the humanities and social sciences. The VABB-SHW was set up in order to be able to give fairer consideration to these disciplines within the model. The VABB must comprise a selection – based on minimum quality requirements – of peer-reviewed academic publications by social sciences and humanities researchers at Flemish universities that are not included in the Web of Science (articles in journals, books and contributions to books, contributions to proceedings) but which must still be considered as output in the distribution model.[45] A special committee for the social sciences and humanities (the 'Authoritative Panel') determines which publications – taken from a complete overview of publications by researchers at the Flemish universities – will be admitted to the VABB and therefore which publications will count within the funding model. The number of types of publications is restricted: articles in journals, books, edited volumes, contributions to books, and articles in proceedings.[46] The classification is binary: publications either count or do not count; there is no further ranking. The database contains only the references to the publications (i.e. their bibliographical specifications) and not the publications themselves; it can therefore not be used to produce citation statistics. It is updated annually and once every three years a review is carried out of the working method and criteria by a group of international experts.

43 For a discussion of the Norwegian model, see Sivertsen (2010) 22–28.

44 For a full description, see Engels et al. (2008).

45 The criteria for admission are comparable to those used in Norway: publicly accessible, identifiable by means of an ISBN or ISSN, contributing to the development of new insights, assessed in a peer review process, attributable to a discipline within the social sciences or humanities.

46 The option has been left open of including other types of product such as editions of texts, catalogues, and databases in a future stage.

This method of classifying publications – on the basis of a classification of journals and publishers – as a quality indicator is open to criticism. Here too, for example, there is not necessarily any one-to-one correspondence between the quality of a publication and the status (for example A, B, or C) of the publication medium, in the sense that a publication in a B or C journal may also be one of high quality, all the more so because some of these journals are in fact niche journals that publish virtually all the publications on a particular subject. The Committee concludes, nevertheless, that classifications of this kind can play a useful role in so far as they create clearer overall distinctions in the research output of a researcher or research group: the fact that someone is able to publish fairly regularly in A journals can be seen as a sign of quality. In the view of the Committee, however, allocating a weighting factor to publications in A, B, and C journals and using it within an evaluative points system suggests an exactitude that this specific quality indicator cannot in fact provide. The Committee therefore considers that the classification into A, B, and C journals within a system of quality indicators in research reports and research assessments should indeed play a relevant role but that the weighting involved should not be an automatic calculation and should be left to the expert judgment of peers.

The question of whether one wishes to utilise classifications of this type – whether or not together with scoring – when allocating funding, as is the case in Norway and Flanders, is a more far-reaching and political question, and one that the Committee does not wish to answer. The Committee does consider, however, that precisely where digitised bibliometrics (citation scores, *h*-index, etc.) are insufficiently usable in many areas of the humanities, a system for classifying publication media (both journals and publishers and books series) can provide useful alternative indicators if applied judiciously and with an awareness of the limitations outlined above.

The Committee is of the opinion that for the time being such lists of A, B, and C journals can best be drawn up at a national level, with the initiatives in other countries also functioning as useful benchmarks. The Committee also believes that in the Netherlands a task of this kind should be performed by discipline-specific bodies, in the first instance probably the national research schools. The lists must also be reassessed from time to time by those same bodies. This process of classification will need to be managed, and the Committee believes that the Academy – given its independent position, authority, and involvement with the national research schools – is well suited to taking on the management role. A survey of the possibilities for creating such a classification for a number of selected disciplines will form part of the pilot phase of the project (see Section 5).

Conclusions

Bibliometric quality indicators must be utilised cautiously. Counting articles measures only productivity; counting citations measures impact, which is not necessarily the same thing as quality and is also often dependent on the publication and citation culture within a given discipline or subdiscipline. Both types of counting can only be

effective in those areas that are to a great extent covered by the Web of Science and other relevant databases. In the case of the humanities, the situation is still highly unsatisfactory: too many book publications and articles in languages other than English are still left out of consideration, meaning that citation statistics mean very little. Once the databases used for certain disciplines in the humanities have achieved a sufficient level of coverage, the use of bibliometric indicators will become more relevant. But even then it will be necessary to guard against simplistic ideas regarding their usefulness. Bibliometric indicators must always be viewed in connection with other indicators and within the qualitative context of peer review.

As an alternative, the Committee supports the development of lists of A, B, and C journals for various disciplines as a means of achieving a minimum level of comparability between various specialisations. Here too, however, it is necessary to stress that such lists will need to be utilised judiciously and within the framework of a balanced peer review procedure. Lists of journals will also need to be dynamic. They will apply for only a limited period of time and must be reviewed periodically by an authoritative body.

CONDITIONS

The considerations in this section produce two new conditions for quality indicators:

- Bibliometric quality indicators must be utilised cautiously. Where they are applicable at all, they must operate within a system of research assessment via peer review: weighting and interpretation must be carried out by peers and other experts and cannot be simply a question of automatic calculation.
- A system of quality indicators for the humanities must be able to rely on a broadly accepted ranking of publication media (journals and books or series).

5 TOWARDS A SYSTEM OF QUALITY INDICATORS FOR THE HUMANITIES

This final section first sets out the conditions for an appropriate system of quality indicators for the humanities. These conditions are based on the conclusions drawn in the preceding sections. A large number of them are then incorporated into a sketch of a system of quality indicators. Finally, the remaining conditions are incorporated into recommendations for the application and implementation of this system, and into guidelines for the pilot phase of this project.

The necessary conditions

The preceding sections clarified the overall context within which a system of quality indicators ought to function. It was noted that quality indicators are necessary in order to reinforce the process of peer review and provide it with an external basis (Section 1). A *system* of quality indicators should also simplify the process of peer review by structuring it (Section 3). Such a system must comprise both output indicators and esteem indicators and indicators of both scholarly and societal quality. The system needs to be as broad as scholarly practice demands, but at the same time sufficiently flexible to enable a tailor-made approach in various contexts. It must also offer scope for additions and adjustments if such are required by the further digitisation of scholarly production processes (Section 1).

We also asserted that compared to other fields of science the humanities do not require a different approach to quality and that they therefore do not require a different *type* of quality indicators. They do, however, demand a fairly wide range of quality indicators that does justice to the diversity of products, target groups, and publishing

cultures that are present within this field; in this, however, they are no different from most other fields of research. This means that primacy cannot be allocated to bibliometric indicators based on databases made up primarily of English-language journal publications. The system must also offer scope for other types of output than just scholarly publications, for example databases, catalogues, and editions of texts (Section 2).

We also stressed that peer review and indicators interact with one another: indicators support peer review, but the specific weighting of the various indicators requires the expert judgment of the peers. In the current circumstances, this applies particularly – but certainly not exclusively – to bibliometric indicators. Isolated and ill-considered use of bibliometric indicators is irresponsible in general and particularly so in the humanities. The Committee also considers, finally, that in order to create clearer distinctions in the output indicators, a national classification is required of journals, publishers, and books series in the humanities. However, this classification too should act mainly as input for the process of qualitative peer review and not as a basis for automatic calculation (Sections 3 and 4).

Some stakeholders adopt the view that a simple system – i.e. one with a few basic indicators that greatly objectifies or even replaces qualitative appraisal by peers – is desirable and feasible. In the view of the Committee, that is not the case. An effective system is not simple and a limited system is not effective. The Committee has, however, sought a system that is as simple as possible – i.e. practicable – because of its flexibility, and in its recommendations for implementation it has attempted at all times to consider the aim of avoiding an extra bureaucratic burden (Section 3).

Sketch of a system of quality indicators for the humanities

The Committee envisages a system of quality indicators for the humanities that takes account of the above considerations and conditions as follows:

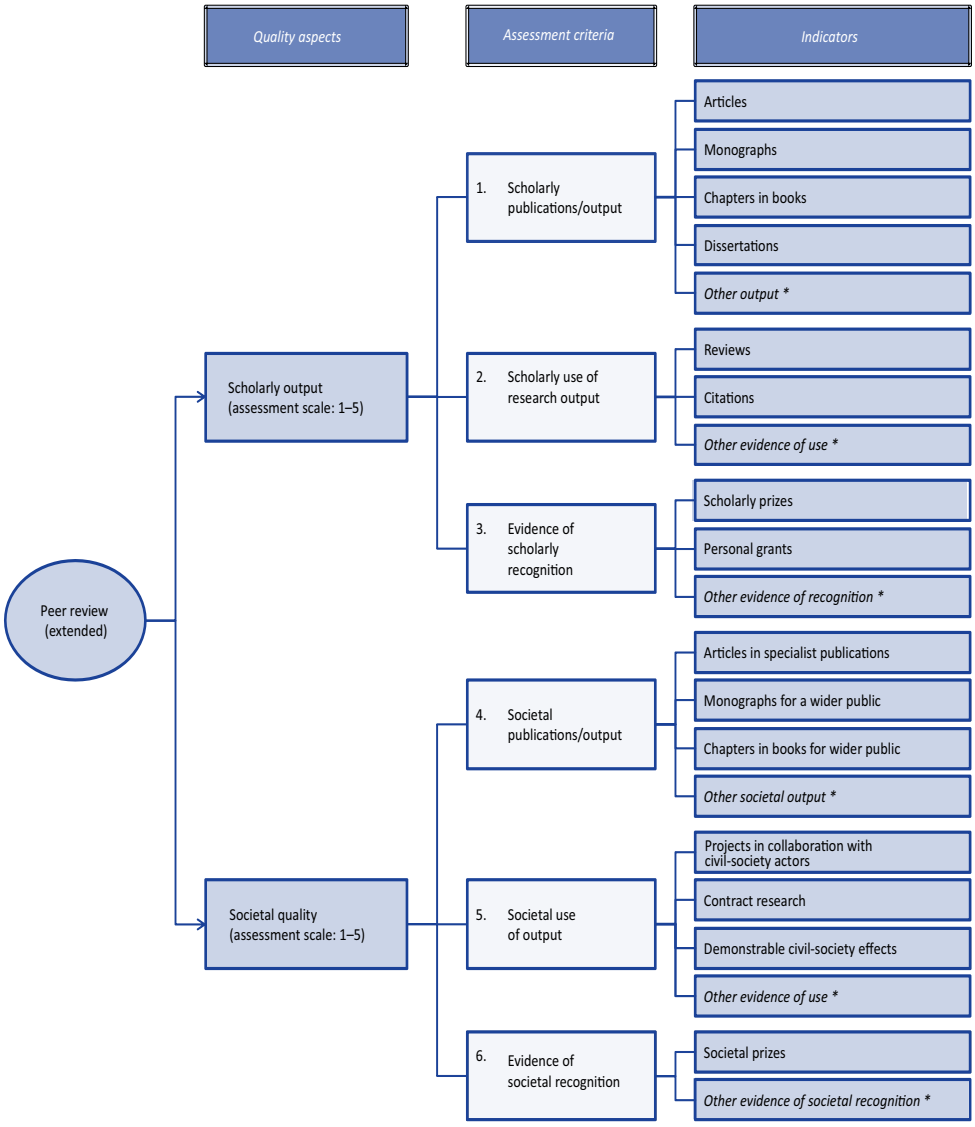

Figure 1 Diagram of criteria and indicators
* discipline-specific/context-specific indicators

A number of aspects of this system and the way in which it should be applied require explanation. Taking the diagram from left to right:

- In assessing research in the humanities, one can distinguish between scholarly quality and societal quality. The Committee proposes that peers, if necessary assisted by external experts in the field of societal quality or societal stakeholders,[47]

47 For this type of extended peer review, see Section 3; see also the remarks below regarding the pilot phase of this project.

should carry out a standard assessment of both aspects according to a five-point scale in line with the Standard Evaluation Protocol.
- Peers, if necessary assisted by external experts in the field of societal quality or societal stakeholders, should assess these aspects on the basis of the assessment criteria specified in the second column. For both scientific and societal quality, these are output, demonstrable use of output, and evidence of recognition ('esteem'). These assessment criteria are what the peers must express their views about. Taken together, these assessments produce the final assessment (on a five-point scale) for the two main aspects (left-hand column).
- The right-hand column lists indicators according to criterion. The disciplines, institutions, institutes, or programmes that are to be subjected to review will determine for themselves the indicators that they do or do not utilise (i.e. those that are relevant to the specific discipline or relevant in connection with the specific mission of an institute or programme or the specific duties of individual researchers).[48] The Committee has not specified any basic set of indicators because a prescriptive demarcation of what that basic set ought to comprise would in any case be arbitrary and one can expect that in practice it will become clear automatically which indicators – for example scholarly articles and monographs – will be considered relevant in the great majority of contexts.
- Guidelines will need to be drawn up regarding the provision of the relevant data. In the case of scholarly articles, for example, the following might be required: complete bibliographical details, numbers in A, B, and C media, and a specific set of sample publications. The Committee wishes to use the pilot phase of this project to draw up practicable guidelines. An initial diagrammatic design (which can act as input for the pilot phase) can be found in Appendix 1.

This system of quality indicators meets the conditions specified in the preceding sections. It comprises both scholarly and societal quality and both output indicators and esteem indicators. It is both broad and flexible: it makes it possible, where appropriate, to omit certain indicators (for example bibliometric indicators) and the category 'other indicators' can be filled in for each of the six assessment criteria in a subject-specific or context-specific manner. The addition of new indicators on the basis of advancing digitisation of scholarly production processes is also possible. The system is therefore not set in stone. It also allows for a quantitative design of some indicators (numbers of publications, key bibliometric figures) but the ultimate appraisal of these criteria in order to arrive at a final judgment of the quality aspects (left-hand column) is a qualitative assessment and will be carried out by peers.

48 By emphasising the importance of creating scope for mission-related quality indicators, the Committee endorses the conclusions in the Academy's report *Judging Research on its Merits* (2005).

Contexts of application

As pointed out in Section 1, quality assessment takes place in different contexts and with different intentions. We deal briefly below with the way in which the above diagram could function in three standard contexts: (a) periodical external reviews, (b) assessment of research proposals and individual CVs, and (c) assessment of the performance of individual staff members and faculties and institutes.

(a) External reviews of research in the framework of the SEP protocol

Publicly financed research institutes and their research programmes are subjected to external review once every six years. For the purpose of such review, the VSNU, NWO, and the Academy have drawn up the Standard Evaluation Protocol (SEP). Some of the quality criteria that the protocol contains – originality, the cohesiveness of a programme (see criterion A1 of the SEP protocol) – can only be assessed directly by peers. Where a number of other criteria for scientific/quality are concerned – contribution to the field, output, relevance – but also such criteria as academic reputation (see criterion A4 of the SEP protocol), productivity (see criterion B2), and societal relevance (see criterion C), peer review can be supported by a relevant selection from the system of quality indicators. Indicators for the societal relevance of research can if necessary be refined further on the basis of the pointers that were recently published in the context of the ERiC project.[49]

At the moment, there is not yet any uniform practice of national, discipline-specific external review in the humanities. Where discipline-specific external review does take place, an authoritative body (i.e. a consultative body for the discipline or a national research school) can carry out a minimum selection as considered necessary. If that is not the case, the institute issuing the assignment for external review will do so. Individual institutes or programmes can also specify mission-related indicators.

The Committee believes that the above-mentioned national classification of publication media is indispensable in order to give substance to the system for categorising journals referred to in the SEP into a top 10% and a top 25% (A and B categories).

(b) Assessment of research proposals and individual CVs

In the case of research proposals submitted to the NWO and similar organisations, referees and jury members are expected to give their assessment of both the quality of the proposal and the scientific/scholarly quality of the party submitting it.[50] The

49 ERiC (2010)

50 The assessment framework for requests within the context of the NWO's Innovational Research Incentive Scheme, for example, comprises three assessment aspects: assessment of quality of the researcher, innovative character of the proposal, and assessment of quality of the research proposal.

former aspect is prospective, and can therefore only be assessed directly by peers (referees); the latter aspect is retrospective and in fact amounts to an assessment of the applicant's research CV. The system of quality indicators that we have outlined can serve here as a frame of reference. The same also applies to contexts in which grants are allocated that constitute a prize or recognition of scientific/scholarly excellence (for example the Spinoza Prizes and the Heineken Prizes). In the context of assessments of this kind it is primarily esteem indicators that will play a role, besides the basic indicators.

A problem noted by the NWO and other funding bodies is the poor level of comparability between different disciplines, meaning that the allocation of funds to those disciplines appears to become somewhat arbitrary. That problem could basically be solved by utilising one or more quality indicators as a common yardstick. The Committee considers that such a solution is inappropriate. Although, as we saw in Section 2, various research fields apply more or less the same idea of quality, the diversity of publishing cultures requires a variety of discipline-specific quality indicators. A simple, limited system that makes easy comparisons possible is an illusion.

That does not mean, however, that no compatibility whatsoever is possible between disciplines where quality assessment is concerned. A certain measure of comparability within the humanities can be achieved, for example by working across the board with classifications of journals and book series, as recommended in the present report. In addition, one could also eventually work with bibliometric indicators that have been standardised for each discipline. This currently appears to be premature, however, and it is in any case laborious and still rather crude because publishing and citation cultures can also differ considerably within disciplines.

(c) Performance assessment within faculties and institutes

At practically all faculties and institutes for the humanities, research groups and researchers are assessed annually as regards their scholarly quality and productivity. This is often, but not exclusively, in the form of 'performance and development' interviews in the context of personnel policy. The focus here is generally on assessing research output in relation to the available research time. This can have consequences for the allocation or reallocation of research capacity and therefore for the duties of the individual member of staff. The system of quality indicators can be applied here too, although evaluation in these cases is of course less wide-ranging than in the above-mentioned contexts (a) and (b). In general, it will be possible to restrict oneself to a number of output indicators. An increasing number of humanities faculties and research institutes apply a points system, or are considering introducing such a system, to structure the process of staff appraisal and the allocation of research time. At most humanities faculties, researchers are required to produce an average of at least one article in a scholarly journal annually over a number of years for every 0.2 FTE of research effort. Depending on the context, books are assigned a value equivalent to a

number of journal articles. The previously mentioned national classification of journals, book publishers, and book series will assist in acquiring an overall picture of the arena in which an individual researcher appears.

Follow-up action and recommendations

The analyses in this report and the associated sketch of a system of quality indicators for the humanities mean that the 'conceptual phase' of the Committee's work as outlined in the introduction has been completed. As already stated, that phase will be followed by an experimental 'pilot phase' in which the system's practicality will be investigated, with changes and refinements perhaps being made, and with recommendations being formulated for implementation. That pilot phase within the procedure leading to the final report will comprise the following elements:

(1) At the behest of the Executive Board of Erasmus University Rotterdam (EUR), the CWTS has commenced a study of the publishing culture within the discipline of history. To that end, the METIS output data from the Dutch universities have been requested. Bibliometrics experts will examine the extent of ISI coverage within that data and, in consultation with historians at the EUR, will then determine whether and how the remaining 'gap' can be filled and whether a more or less complete database can be created that can serve as the basis for bibliometric study. The Committee will monitor this project in close consultation with the CWTS with a view to clarifying the possibilities for carrying out bibliometric investigation within a discipline using improved methods. The Committee's approach will be a neutral one, and it is neither committing itself in advance to the usefulness or necessity of such an operation nor to the policy objectives that the EUR has or may have in this regard. The central focus for the Committee is the factual question of what this exercise produces, whether it deserves to be imitated, and whether the added value that may be generated is in proportion to the necessary investment in terms of manpower and money.

(2) The Committee on Quality Indicators in the Humanities will be in charge of two specific pilot studies. These will involve experimental evaluations of the Academy's Meertens Institute and Groningen University's Research Institute for the Study of Culture (ICOG). Both these pilot projects will be used to test the system of quality indicators that has been developed here and will consider specifically

- whether the system can differentiate sufficiently in actual practice;
- what the possibilities and problems are as regards drawing up discipline-specific classifications of publication channels;
- how a system – producing a maximum of information but involving a minimum burden – can be developed for the provision of material for research assessments (for example a modular system that can be used in various different evaluation contexts), and how this relates to the categories within the existing METIS

registration system;
- what further refinement of the indicators and the assessment procedure is or is not possible as regards the aspect of societal quality (for example by involving civil-society stakeholders in the process). The 2010 recommendations in the *Pointers for Evaluating the Societal relevance of Scientific Research* resulting from the ERiC project (Evaluating Research in Context) will be followed.

Findings from the pilot phase will be incorporated into the Committee's final report. In anticipation of that final report, the Committee believes that, on the basis of the present draft report, it can already indicate the outlines of two recommendations to the Academy:

- In the first place, the Academy should take the lead – and then take charge of – a national initiative to arrive at a broadly supported discipline-oriented classification of scholarly journals and book series for the humanities. It would seem obvious that the national research schools should be involved in the first instance; they can, after all, be considered as expert platforms for their respective disciplines.
- Secondly, the Committee notes that its findings agree on a number of crucial points with the recently published report on Quality Assessment in the Design and Construction Disciplines. This makes it opportune to ask what the situation is in other areas of science and scholarship. The Academy could arrange for discussion of quality indicators to take place in its advisory councils for the various research fields so as to survey varying views regarding such indicators and to investigate the possibilities regarding further consensus-forming.

SOURCES

Adema, J. & P. Rutten (2010), *Digital Monographs in the Humanities and Social Sciences: Report on User Needs* (ECP-2007-DILI-537002). Open Access Publishing in European Networks (OAPEN). OAPEN Report.

Arts & Humanities Research Council (2009), *Leading the World. The Economic Impact of UK Arts and Humanities Research.*

AWT (2005), *De waarde van weten. De economische betekenis van universitair onderzoek*. The Hague: AWT.

AWT (2007), *Alfa en gamma stralen. Valorisatiebeleid voor de alfa- en gammawetenschappen*, The Hague: AWT.

Borgman, C. (2007), *Scholarship in the Digital Age: Information, Infrastructure and the Internet,* Boston: MIT Press.

Broek, S.D. & A.J. Nijssen (2009), *Impact assessment geesteswetenschappen*, onderzoeksrapport Research voor Beleid in opdracht van NWO Gebiedsbestuur GW.

Commissie Nationaal Plan Toekomst Geesteswetenschappen (2008), *Duurzame Geesteswetenschappen*, Amsterdam: Amsterdam University Press.

Costas, R. & M. Bordons (2008), 'Is G-Index Better than H-Index? An Exploratory Study at the Individual Level', *Scientometrics* 77, 2.

De Solla Price, D. (1963), *Little Science, Big Science*, New York: Columbia University Press.

Dolan, C. (2007), *Feasibility Study: the Evaluation and Benchmarking of Humanities Research in Europe.* Report of the AHRC for the Humanities in the European Research Area.

Engels, T.C.E., E.H.J. Spruyt, W. Glänzel & K. Debackere (2008), 'Het Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen: instrument ten dienste van een optimaal wetenschapsbeleid?', *TORB*, 2008-09/5.

ERiC (2010), *Handreiking evaluatie van maatschappelijke relevantie van wetenschappelijk onderzoek*, www.eric-project.nl

European Reference Index for the Humanities (ERIH), on line, http://www.esf.org/research-areas/humanities/erih-european-reference-index-for-the-humanities.html

Harzing, A-W. (2008), *Google Scholar – a New Data Source for Citation Analysis*, http://www.harzing.com/pop_gs.htm

Head, S. (2011), 'The Grim Threat to British Universities', *The New York Review of Books*, vol. LVIII, 1 (Jan./Feb. 2011), 58-64.

Hemlin, S. (2009), 'Peer Review Agreement or Peer Review Disagreement: Which is Better?', *Journal of Psychology of Science and Technology*, Vol. 2, No. 1.

Hicks, D. & J. Wang (2009), *Towards a Bibliometric Database for the Social Sciences and Humanities – A European Scoping Project*. School of Public Policy, Georgia Institute of Technology.

Jin & al. (2007), 'The R and AR-indices: Complementing the H-index', *Chinese Science Bulletin* 52, 6.

KNAW (2005), *Judging Research on its Merits*, Rapport van de Raad voor de Geesteswetenschappen en de Sociaal-Wetenschappelijke Raad.

KNAW-commissie Kwaliteitszorg (2008), *Academy Committee for Quality Assurance (2008), Quality Assurance in scientific research: From SEP to CEP: Balancing fairness simplicity.*

Linmans, A.J.M. (2008), *Wetenschappelijk onderzoek in de Faculteit der Letteren van de Universiteit Leiden. Een onderzoek naar bibliometrische indicatoren voor het bepalen van impact van wetenschappelijk onderzoek in de geesteswetenschappen*, Leiden: CWTS.

Merton, R.K., (1968), 'The Matthew Effect in Science', *Science*, 5 January 1968, vol. 159 no. 3810.

Moed, H. F. (2005), *Citation Analysis in Research Evaluation*, Dordrecht: Springer.

Moed, H. F., J. Linmans, A. Nederhof & A. Zuccala, C.L. Illescas & F. de Moya Aragon (2009), *Options for a Comprehensive Database of Research*, Research Report to the Project Board of the Scoping Study 'Towards a Bibliometric Database for the Social Sciences and the Humanities' set up by the Standing Committees for the Social Sciences and the Humanities of the European Science Foundation (ESF). Leiden: CWTS, Madrid/Granada: SCIMago Research Group.

Norwegian Association of Higher Education Institutions (2004), *A Bibliometric Model for Performance-Based Budgeting of Research Institutions* (English translation).

Project Board Bibliometric Database SSH (2010), *Towards a Bibliometric Database for the Social Sciences and Humanities*. Report produced for DFG, ESRC, AHRC, NWO, ANR & ESF.

Sense About Science (2004), *Peer Review and the Acceptance of New Scientific Ideas.* Discussion paper from a working party on equipping the public with an understanding of peer review.

Sivertsen, G. (2010), 'A Performance Indicator Based on Complete Data for the Scientific Publication Output at Research Institutions', *ISSI Newsletter*, Vol. 6, No. 1.

Verhaar, P., M. Mitova, P. Rutten, A. van der Weel, F. Birnie, A. Wagenaar & J. Gloerich (2010), *Data Curation in Arts and Media Research*. Utrecht: SURFfoundation.

VSNU, KNAW and NWO (2009), *Standard Evaluation Protocol 2009-2015,* www.knaw.nl/sep

Waltman, L. & N.J. van Eck (2009), 'A simple alternative to the H-index', *ISSI-Newsletter*, vol. 05 / no. 3, September 2009.

APPENDIX 1

OVERVIEW OF INDICATORS (INPUT FOR THE PILOT PHASE)

Assessment criteria	Indicators	Information requirement (to be made more specific in the pilot phase)
1. Scientific/scholarly publications	Articles	• List • Numbers, marked according to A, B, and C journals • Selection of key publications
	Monographs	• List • Numbers, marked according to A, B, and C publishers and series • Selection of key publications
	Chapters in books	• List • Numbers, marked according to A, B, and C publishers and series • Selection of key publications
	Completed dissertations	• List per research group • List per researcher (main supervisor)
	Other scientific/scholarly output, for example, conference papers, editorships, handbooks, dictionaries, editions of texts, databases, software	Quantitative and/or qualitative information to be delivered as determined according to the context
2. Scientific/scholarly use of research output	Reviews	Traceable references to samples
	Bibliometric indicators	Citation analysis (only if the databases provide sufficient coverage)
	Other evidence of use, for example books sold/loaned, articles downloaded, use of databases, software	Quantitative and/or qualitative information to be delivered as determined according to the context
3. Evidence of scientific/scholarly recognition	Scientific/scholarly prizes	Simple statement with dates (years)
	Personal grants (VENI, VIDI, VICI, Spinoza, ERC)	Simple statement with dates (years)
	Other evidence of scientific/scholarly recognition, for example memberships of academies, honorary doctorates, external appointments, invitations to give keynote lectures, invitations to contribute to handbooks, dictionaries or editions of texts, chief/full editorships of noted national/international journals/ books/collections, advisory/ referee appointments (for example NWO, external review committees), fellowships at prestigious institutes/foreign institutes	Quantitative and/or qualitative information to be delivered as determined according to the context

4. Civil-society publications	Articles in specialist publications (not being primarily scientific/scholarly journals)	• List • Selection of key publications
	Monographs for non-scientists/scholars and interested individuals	• List • Selection of key publications
	Chapters in books for non-scientists/scholars and interested individuals	• List • Selection of key publications
	Other civil-society output, for example collections for non-scientists/scholars and interested individuals, editorships of specialist publications, handbooks, dictionaries, editions of texts, databases, software, exhibitions, catalogues, translations, advisory reports on policy	Quantitative and/or qualitative information to be requested as determined according to the context
5. Civil-society use of research output	Projects carried out in collaboration with civil-society actors	Simple statement with dates (years)
	Contract research	Simple statement with dates (years)
	Demonstrable civil-society effects of research	Simple statement with dates (years)
	Other types of civil-society use, for example reviews, citations in policy reports, use of publications, media attention, books sold/loaned	Quantitative and/or qualitative information to be requested as determined according to the context
6. Evidence of civil-society recognition	Civil-society prizes	Simple statement with dates (years)
	Other evidence of civil-society recognition, for example civil-society appointments, invitations to give lectures, invitations for media appearances, advisory positions/membership of advisory committees	Quantitative and/or qualitative information to be requested as determined according to the context

APPENDIX 2

SOME EXAMPLES OF DEVELOPMENTS REGARDING INDICATORS FOR THE HUMANITIES AND SOCIAL SCIENCES IN OTHER COUNTRIES

The use of indicators in the humanities and social sciences is a topic of discussion in many countries. The report *Judging Research on its Merits* produced by the SWR and the RGW in 2005 is an example in the Dutch context. It does not propose a system but it does indicate the direction in which a new system of indicators could develop. In the present report, we have dealt at length with two examples of systems that are in an advanced stage of development (in Norway and Flanders). We will not deal with them any further here. We give a number of examples below of studies and proposals that in part underlie the thinking within the Committee. These do not involve fully worked-out systems but they do indicate the main points of discussion for a number of the new indicators that are to be developed. The indicators concerned are for both assessment of scientific/scholarly quality and societal impact. We give a short presentation of each study, followed by the full summary taken over verbatim from the reports concerned.

1. Royal Flemish Academy of Belgium for Science and the Arts (Brussels) [*Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten*], *Bibliometrie in de Humane Wetenschappen* [Bibliometrics in the Humanities] (2004), report by the working party on bibliometrics in the humanities.

This is a report by two committees whose remit was to investigate the possibility of a bibliometric system for the humanities and social sciences. It can be seen as a precursor to the Flemish Academic Bibliographic Database discussed in the present report.

Summary

The KVAB and KANTL committees were set up to investigate new options for the development of bibliometric systems for humanities and social science research, aiming at a more objective comparison and enhancing the quality of research. They do not underestimate the importance, especially for young scholars, of escaping parochialism, and of having one's research judged by international standards in the field. However, after careful investigation of the new developments, they feel obliged to warn of certain unintended consequences of these new policies, which may turn out to be

counterproductive, even detrimental to the quality of research – particularly in some (parts) of the Human Science disciplines. Ultimately, the question is whether bibliometry will be in the service of research or whether research will be done solely as a function of obtaining career-furthering bibliometrical results. (That these worries are also present in the Exact Sciences perhaps escapes certain policy makers; see: Peter A. Lawrence, 'The politics of publication', *Nature* Vol. 422, 20 March 2003, p. 259-261.)

What are some of the major worries of the committees? In the first place, there is the excessive emphasis of present proposals on publication in journals (particularly English language journals). This runs counter to such facts as that books may be as important as, or even more important than articles in certain disciplines, and that, in some disciplines, the 'forum' language may rather be French, Italian, or even Dutch, etc. To suppose that research can always be reorganized so as to fit into the mould of publications of supposed international standing, is to forgo

- the very diverse nature of disciplines in the Human Sciences;
- the absence, sometimes, of a generally accepted methodology or paradigm;
- the desirability, in some cases, of an essayistic presentation of results;
- and the intrinsic link between certain (parts of) disciplines and an orientation towards 'conversations' in the surrounding culture or society, etc.

However difficult the evaluation of publications may be in the Human Sciences, it would be absurd to think that, in view of the lack up to now of certain measuring tools, judgments about quality of research were absent from the Human Sciences. What is needed is a careful investigation of existing practices of adequate (?) evaluation both in university nomination committees and research councils, and in national and supranational research evaluation committees (there is clearly a need here for more research in this matter, research which inevitably will be the task of certain Human Science disciplines). Bibliometry is of course not to be excluded in evaluation, but it should play a subordinate role, the role of a preliminary test, which should be taken seriously, but which has to be supplemented with other considerations. A good deal of work remains to be done to come up with tools which are sufficiently sophisticated for measuring quality in a valid and reliable way in the fields concerned. Scholars working in the Human Sciences have the collective responsibility to ensure that these tools are really adapted to their task.

Some entertain the wish to establish bibliometrical systems which would allow an objective comparison between the quantity and quality of output of the most diverse researchers, or research groups, in whatever scientific disciplines. This wish, unjustly, supposes that research output is fundamentally similar in the different scientific disciplines. Unfortunately (?), this is not even the case in the Positive Sciences. Therefore, other ways have to be found to make comparisons and decisions concerning vastly different kinds of research output.

2. ESF project, *Towards a Bibliometric Database for the Social Sciences and Humanities*

– *A European Scoping Project*, a report produced for DFG, ESRC, AHRC, NWO, ANR, and ESF, 8 March 2010

This was a wide-ranging study set up by a number of research councils. It was submitted to the parties that commissioned it in mid-2010 but we are not aware of any follow-up. This study ran parallel with an attempt to produce a European Reference Index for the Humanities (ERIH), but was in fact separate from it.

Executive Summary

The aim of the project was to explore the possibility of developing a database for capturing the full range of research outputs from the Social Sciences and Humanities (SSH). SSH research outputs include not just those articles published in international journals, but also articles in national journals, academic book chapters and books, books aimed at a more popular audience, monographs, reports in the 'grey literature', and non-published outputs from fields such as the performing arts. Such a database is intended to serve a number of related purposes. One is to address the growing pressure from policy-makers and research funders to demonstrate 'accountability' and to ensure 'value for money'.

A second is to develop performance measures for assessing research quality and impact. For basic research in the natural sciences, there are fairly well established indicators of research output and impact (based on publications and citations), but these work poorly for the social sciences and barely at all for arts and humanities.

Third, research funders and others may want to use the database to provide an overview of SSH research outputs in Europe.

Fourth, funders and policy makers may use it as a source of information or to identify areas of strong research capacity and those that are perhaps in need of capacity-building or support.

Lastly, research councils may seek to use the database as a tool for mapping emerging areas of (often interdisciplinary) research. The question addressed in this project is, 'What is the potential for developing some form of research output database that could be used for assessing research performance in SSH?' An important aspect to the background context of the study relates to recent developments with respect to various databases and bibliographic lists on which a SSH bibliometric database might build. These include the growth of 'Open Access' publications, significantly improved coverage of journal literature by the Web of Science and Scopus, the emergence of Google Scholar as a new source of publication and citation data for books as well as journals, the growing availability of digital data on publications from book publishers, the development of various national or disciplinary bibliographic databases, and the establishment of a range of institutional repositories for research outputs. The report identifies the main problems and issues to be confronted in any attempt to construct an inclusive SSH bibliometric database. It analyses a number of key considerations for the creation of an inclusive SSH bibliometric database, along with certain operational

issues. On this basis, it set outs various strategic options. The report concludes with a number of recommendations chosen on the basis of their practicality and cost-effectiveness. We propose a way forward based on four main recommendations.

1. The first centres on the definition of criteria as to which SSH research outputs should be included in a bibliometric database and the establishment of a standardised database structure for national bibliometric databases.
2. The second explores the option of involving a commercial supplier in the construction of a single international SSH bibliometric database. In both of these, the focus is on published scholarly outputs appearing either in peer-reviewed journals, or in books that have likewise been subject to peer-review before publication.
3. The third recommendation involves conducting a small pilot study focusing on one or more specific SSH disciplines.
4. The fourth recommendation deals with the longer-term expansion and enhancement of the SSH bibliometric database to include other SSH outputs (that is, in addition to peer-reviewed articles and books).

A hybrid approach (i.e. a combination of 'top-down' and 'bottom-up') has been recommended for the implementation of each recommendation because this appears to offer the best of both worlds – impetus, guidance and authority from the top, and expertise and experience from the bottom. The report argues that these four recommendations offer the most promising way forward in exploring and then establishing an international bibliometric database for the social sciences and humanities.

3. Carmen López Illescas, Felix de Moya Anegón, Janus Linmans, Anton Nederhof and Henk F. Moed, *Options for a comprehensive database of research outputs in social sciences and humanities* (2009).

This is a joint project by the SCIMago Research Group, CSIC Madrid, University of Granada, Spain, and the Centre for Science and Technology Studies (CWTS), Leiden University. It investigates the content of a large number of bibliographical and bibliometric databases and their suitability for use in the social sciences and humanities (they include Web of Science, Google and Google Scholar, and Scopus).

Executive summary

This report presents the outcomes of a study aimed at:

- Examining the potentialities of current bibliographic databases of scientific-scholarly literature as sources of indicators of research performance in social sciences and humanities (SSH);
- Proposing options for the creation of a comprehensive database of research outputs in these domains of scholarship;
- Highlighting actual practices in using bibliometric indicators in SSH fields;

- Exploring means of capturing and assessing non-published outputs.

Recommendations:

1. Combine a number of existing European special SSH bibliographies;
2. Create a new database of SSH outputs from publishers' archives;
3. Stimulate further enhancement of SSH coverage of Web of Science and /or Scopus;
4. Stimulate further development of institutional repositories;
5. Stimulate creation and standardization of institutional research management systems;
6. Explore the potentialities and limitations of Google Scholar and Google Book Search.

4. Michèle Dassa, Christine Kosmopoulos and Denise Pumain, *JournalBase – A Comparative International Study of Scientific Journal Databases in the Social Sciences and the Humanities (SSH)*, Cybergeo, European journal of Geography, no. 484, 2010

This is a comparative study of the coverage of the best-known databases for the humanities and social sciences.

Abstract

Presented for the first time, a comparative table of the contents of the databases that list the journals in the Social Sciences and the Humanities (SSH) named *Journal Base* has been published, in June 2009, in *Cybergeo.* The study covers the AHCI and the SSCI of the *Web of Science* published by Thomson Reuters) and *Scopus* (published by Elsevier), as well as the *European Reference Index for Humanities* (ERIH) (published by the *European Science Foundation*) and the lists of the French *Agence p'ur l+Evaluation de la Recherc'e et de l+Enseignement Supérieur* (AERES). The research was carried out in 2008 with the financial support of the TGE Adonis of the CNRS. With some 20,000 entries, corresponding to 10,000 different journals, this is an almost exhaustive overview of the wealth of publications in the Social Sciences and the Humanities, at last made available in this table, adopting the same nomenclature for classifying the journals according to their disciplines as the one used in 27 headlines of the *European Science Foundation*. The multiple assignments reveal the multidisciplinarity of the journals, which is quite frequent in SSH, but also sometimes the inconsistency of databases that have not been corrected. The article presents the history of the project, the methodology established by the authors, the difficulties encountered in comparing the data. The first results reveal a broader coverage of the ERIH list for the human sciences and an over-representation of the Anglophone journals in the commercial databases throughout the disciplines. The objective of this work is to provide information about the contents of these databases, at a time when bibliometric tools are raising much discussion about their application in the field of social sciences and humanities.

5. Howard D. White, Sebastian K. Boell, Hairong Yu,'Mari Davis, Concepción S. Wilson, and Fletcher T. H. Cole, *Libcitations: A Measure for Comparative Assessment of Book Publications in the 'umanities and Social Sciences+, *Journal of the American Society for Information Science and Technology*, 60(6):1083–1096, 2009

This is a proposal for developing a measure for the assessment of books. The measure proposed is based on the choices made by librarians, on the assumption that those choices are based on a knowledge of the reputations of researchers and of the interests of potential readers.

Abstract
Bibliometric measures for evaluating research units in the book-oriented humanities and social sciences are underdeveloped relative to those available for journal oriented science and technology. We therefore present a new measure designed for book-oriented fields: the 'libcitation count.' This is a count of the libraries holding a given book, as reported in a national or international union catalogue. As librarians decide what to acquire for the audiences they serve, they jointly constitute an instrument for gauging the cultural impact of books. Their decisions are informed by knowledge not only of audiences but also of the book world (e.g., the reputations of authors and the prestige of publishers). From libcitation counts, measures can be derived for comparing research units. Here, we imagine a match-up between the departments of history, philosophy, and political science at the University of New South Wales and the University of Sydney in Australia. We chose the 12 books from each department that had the highest libcitation counts in the Libraries Australia union catalog during 2000 to 2006. We present each book's raw libcitation count, its rank within its Library of Congress (LC) class, and its LCclass normalized libcitation score. The latter is patterned on the item-oriented field normalized citation score used in evaluative bibliometrics. Summary statistics based on these measures allow the departments to be compared for cultural impact. Our work has implications for programs such as Excellence in Research for Australia and the Research Assessment Exercise in the United Kingdom. It also has implications for data mining in OCLC's WorldCat.

6. The British Arts and Humanities Research Council (AHRC) has been working for many years to demonstrate the impact of research. Two examples are given below of the impact of artworks (in this case digital) on other disciplines.

In recent years there has been mounting concern to understand the distribution, utility and influence of research findings in non-academic contexts. This concern originates in part from political imperatives to demonstrate public value, for research to move towards pragmatic considerations in wider public discourse, in cultural, industry and policy environments.

All UK Research Councils are expected to be able to demonstrate the wider impact and value of academic research. The important question that we must seek to address is: what is the contribution of arts and humanities research to society? Or, what is the impact or influence of arts and humanities research outside the academy? The Arts and Humanities Research Council has commissioned a series of case studies to investigate the impact of arts and humanities research. Across the series as a whole, impact has been defined in its broadest sense to include economic, social and cultural elements. The case studies included in this publication focus on the social impact of two artist exhibitions, specifically concentrating on visitor responses and reactions.

Translations, *by Jim Pattison*
A series of digital artworks interpreting the experiences and language of dialysis and kidney transplantation. *Translations* shows how art can be an important medium in the communication of medical terminology between practitioners and patients, and how it can help scientists to innovate by looking beyond the aesthetic constructs that are taken for granted in images. It also gives insights into medical conditions.

In Search of a Hidden Landscape, *by David Walker Barker.*
Artworks about the end of lead and fluorspar mining in the Northern Pennines and the interconnections between the human and physical features of the landscape.

In Search of a Hidden Landscape is concerned with how landscapes are formed out of human necessity and how this gives an insight into the complexities of geological and human roots. It also encourages visitors to value the local links and connections that create and reinforce social capital.

7. The impact Group, Canada, *The Economic Role and Influence of the Social Sciences and Humanities: A Conjecture*, March 2008

This is a Canadian study of the impact of the social sciences and humanities on the economy. It is interesting to see how an indication is given, on the basis of specific data, of how important research in the social sciences and humanities is for the economy.

Summary
This essay explores the role and influence of the social sciences and humanities (SSH) on one aspect of society – the economy. The project has several objectives:
- to develop a framework within which the economic role and influence of the SSH can be examined;
- to stimulate a broader discussion of the economic role and influence of the SSH within the community of SSH scholars and researchers;
- to motivate additional academic research; and to contribute to increased awareness of the role that the SSH – and SSHRC – play in society at large and particularly

in its economic life. Key findings are:

- The UK has recently abandoned its effort to develop a universal algorithm to calculate the economic impact of research, indicating the task is difficult.
- SSH research falls into 29 major disciplinary categories. At least 9 of these have strong links to the economy (Communications and Media Studies, Demography, Economics, Education, Geography, Industrial Relations, Management, Business and Administrative Studies, Urban and Regional Studies, and Law.
- An additional 11 disciplines have moderate links to the economy. A further 9 disciplines have indirect links.
- Of 46 possible areas of application of SSH research, 26 areas have direct relevance to the economy (e.g. Biotechnology, Management) and 30 have indirect relevance (e.g. Children and Youth, Global/Climate Change).
- Service industries now account for 69% of Canada's economic output, whereas good producing industries account for 31%. Service industries rely primarily on SSH for their knowledge inputs while goods-producing industries primarily rely on STEM (science, technology, engineering, medicine). Both groups of industries draw on knowledge from the other sources.
- SSH-based industries account for about 76% of total employment, whereas STEM industries account for about 24%.
- Two-thirds of industries appear to rely primarily on SSH inputs vs. one-third on STEM inputs.
- There are many examples of high-profile companies (and government departments) that make heavy use of SSH inputs to their work or whose business is to produce SSH outputs:
- Thomson Corporation, Torstar Corporation, Cossette Communication Group, Cirque du Soleil, Canadian Western Bank, Manulife Financial Corporation, Stratford Festival, Industry Canada.
- A number of inherently SSH-type processes (e.g. business strategies, management practices, intellectual property protection, organizational changes, improved communications) are essential if firms are to successfully apply knowledge.
- Knowledge can be codified, embodied or tacit. End-user organizations make us of all three types.
- Scientific research influences the economy by: generating concepts and ideas, training people, providing analytical tools, developing technology, or generating specific data and information. SSH and STEM research are equally capable of generating influence.
- Industries that rely primarily on SSH inputs account for $696.7 billion of annual GDP output.
- Industries that rely primarily on STEM inputs account for $431.4 billion of GDP.
- Notionally, among 18 industries: SSH knowledge has a high influence on 7 industry groupings, a moderate influence on 4 industries and a low influence on 7. The

results for STEM are similar.

- By attributing a level of influence for SSH and STEM to each industry, we have inferred that SSH influences $389 billion of economic activity, close to the $400 billion influenced by STEM.

APPENDIX 3

LIST OF PEOPLE INTERVIEWED

Prof. Hans Bertens	Utrecht University
Prof. Wim Blockmans	Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS)/Leiden University
Annemarie Bos	Netherlands Organisation for Scientific Research (NWO)
Prof. Kees de Bot	University of Groningen
Prof. Rosemarie Buikema	Netherlands Graduate School for Women's' Studies
Prof. Karin Bijsterveld	Netherlands Graduate School of Science, Technology and Modern Culture (WTMC)
Liesbeth Cilek	Utrecht University
Prof. Koenraad Debackere	Katholieke Universiteit Leuven
Dr. Hester Dibbits	Meertens Institute
Prof. Martin Everaert	Utrecht Institute of Linguistics OTS
Prof. Joris van Eijnatten	Utrecht University
Prof. Pol Ghesquière	Katholieke Universiteit Leuven
Elisabeth d'Halleweyn	Dutch Language Union
Patricia van Hemert	Vrije Universiteit Amsterdam
Heidi van den Heuvel	University of Groningen
Prof. dr. Frans Hinskens	Meertens Institute
Erna Kas	Utrecht University
Dr Eloe Kingma	Amsterdam School for Cultural Analysis (ASCA)
Paul Koopman	Temporary Task Force for the Humanities
Prof. Anne-Marie Korte	Netherlands School for Advanced Studies in Theology and Religion (NOSTER)
Peter van Laarhoven	Groningen University Library
Dr Janus Linmans	Centre for Science and Technology Studies (CWTS)
Prof. Eric Moorman	Research Institute for Historical, Literary and Cultural Studies
Hotze Mulder	University of Amsterdam
Prof. Peter Nissen	Netherlands School for Advanced Studies in Theology and Religion (NOSTER)
Prof. Koen Ottenheym	Dutch Postgraduate School for Art History

Prof. Frits van Oostrom	Temporary Task Force for the Humanities
Prof. Maarten Prak	Research Institute for History and Culture
Dr Frans Ruiter	Research Institute for History and Culture
Dr Fransisca Snoeck Henkemans	Amsterdam School for Cultural Analysis (ASCA)
Prof. Henriette de Swart	Netherlands National Graduate School of Linguistics
Dr Jeroen Touwen	N.W. Posthumus Institute
Prof. Paul Wouters	Centre for Science and Technology Studies (CWTS)
Prof. Jan Luiten van Zanden	Utrecht University
Prof. Erik-Jan Zürcher	Council for the Humanities

Discussion with the Flemish Authoritative Panel for the Humanities and Social Sciences, August 2010.

Participation in seminar on performance measurement in the humanities, Berlin, January 2010.