

KWALITEITSINDICATOREN VOOR ONDERZOEK  
IN DE GEESTESWETENSCHAPPEN



2011 Koninklijke Nederlandse Akademie van Wetenschappen (KNAW)

© Sommige rechten zijn voorbehouden / Some rights reserved

Voor deze uitgave zijn gebruiksrechten van toepassing zoals vastgelegd in de Creative Commons licentie. [Naamsvermelding 3.0 Nederland]. Voor de volledige tekst van deze licentie zie <http://www.creativecommons.org/licenses/by/3.0/nl/>

Koninklijke Nederlandse Akademie van Wetenschappen

Postbus 19121, 1000 GC Amsterdam

Telefoon + 31 20 551 0700

Fax + 31 20 620 4941

[knaw@bureau.knaw.nl](mailto:knaw@bureau.knaw.nl)

[www.knaw.nl](http://www.knaw.nl)


pdf beschikbaar op [www.knaw.nl](http://www.knaw.nl)

Basisvormgeving: edenspiekermann, Amsterdam

Opmaak: Ellen Bouma, Alkmaar

Illustratie omslag: Dragan85

ISBN: 978-90-6984-625-5

Het papier van deze uitgave voldoet aan  iso-norm 9706 (1994) voor permanent houdbaar papier.

# KWALITEITSINDICATOREN VOOR ONDERZOEK IN DE GEESTESWETENSCHAPPEN

Koninklijke Nederlandse Akademie van Wetenschappen  
Interim-rapport van de Commissie Kwaliteitsindicatoren Geesteswetenschappen  
mei 2011



# VOORWOORD

De schaarste aan middelen voor wetenschappelijk onderzoek en het streven om de beperkte middelen aan kwalitatief hoogwaardig onderzoek te besteden, maken het meten en beoordelen van de kwaliteit van wetenschappelijk onderzoek steeds belangrijker. Dit geldt ook voor de geesteswetenschappen.

Over de juiste methoden voor kwaliteitsbeoordeling bestaat echter vrij breed discussie, ook binnen de geesteswetenschappen. Al in 2005 plaatste de KNAW kritische noten bij de toepassing van bèta-georiënteerde methoden bij kwaliteitsbeoordeling van geesteswetenschappelijk onderzoek. In het rapport *Judging Research on its Merits* pleitte zij toen voor ‘eigen’ meetmethoden voor de geestes- en sociale wetenschappen en voor het meewegen van de maatschappelijke waarde van deze wetenschapsgebieden.

In 2009 stelde de Commissie Nationaal Plan Toekomst Geesteswetenschappen (‘Commissie Cohen’) dat er in de geesteswetenschappen sprake is van een ‘inadequaat instrumentarium van kwaliteitsobjectivering en –differentiatie’. De Commissie Cohen vroeg daarop de KNAW het voortouw te nemen in de ontwikkeling van een maximaal eenvoudig, helder en adequaat systeem van kwaliteitsindicatoren in de Nederlandse geesteswetenschappen. De KNAW is gaarne op dit verzoek ingegaan, mede in het licht van het hier bovengenoemde rapport. Het Dagelijks Bestuur van de KNAW heeft in oktober 2009 de Commissie Kwaliteitsindicatoren Geesteswetenschappen ingesteld met als opdracht om, mede op grond van een veldconsultatie en een internationale inventarisatie, een systeem van kwaliteitsindicatoren voor de Nederlandse geesteswetenschappen te ontwerpen. De werkzaamheid van dit systeem zou door het uitzetten van *pilots* op zijn werkzaamheid empirisch moeten worden getoetst.

Het voorliggende rapport is een interim-rapport waarin de Commissie op grond van zijn bevindingen een voorstel doet dat nu in de praktijk zal worden getoetst. Het is het bestuur van de KNAW opgevallen dat het voorstel van de Commissie

Kwaliteitsindicatoren Geesteswetenschappen belangrijke overeenkomsten vertoont met het onlangs eveneens door de KNAW uitgebrachte advies voor de ontwerpende en construerende wetenschappen. Beide rapporten gaan uit van twee hoofdcriteria voor beoordeling, wetenschappelijke kwaliteit en maatschappelijke relevantie. Beide rapporten stellen indicatoren voor output, gebruik en erkenning voor. En beide rapporten plaatsen peer review centraal in de beoordeling. In het licht van deze overeenkomsten zal de KNAW dan ook bevorderen dat er een bredere discussie zal worden gevoerd over de vraag of deze overeenkomsten kunnen leiden tot een vernieuwde benadering van kwaliteitsbeoordeling in verschillende, of mogelijk alle, vakgebieden. De KNAW heeft samen met het rectorenoverleg en het ERiC-project het initiatief genomen tot een inventarisatie van recente studies op dit gebied. De resultaten hiervan zullen na de zomer bekend worden.

Tegelijkertijd kijkt het bestuur van de KNAW uit naar de resultaten van de twee pilotstudies die bij het Meertens Instituut van de KNAW en bij het Instituut voor Cultuurwetenschappelijk Onderzoek Groningen (ICOG) van de Rijksuniversiteit Groningen zullen worden uitgevoerd. De verwachting bestaat dat beide studies elkaar zullen verrijken en mogelijk bouwstenen kunnen leveren voor een nieuwe benadering van beoordeling waarin de rijkdom van wetenschappelijk werk in uiteenlopende vakgebieden recht wordt gedaan.

Robbert Dijkgraaf  
President KNAW

# INHOUD

VOORWOORD	5
SAMENVATTING	9
TEN GELEIDE	15
1. KWALITEITSBEOORDELING EN KWALITEITSINDICATOREN	19
Kwaliteit van wetenschappelijk onderzoek	19
Kwaliteitsbeoordelingen in verschillende contexten	20
Standaardisering van <i>peer review</i> -procedures	21
Kwaliteitsindicatoren	22
Recente ontwikkelingen	22
Conclusies	25
2. ONDERZOEK EN ONDERZOEKSKWALITEIT IN DE GEESTESWETENSCHAPPEN	27
De eigenheid van de geesteswetenschappen	27
Diversiteit binnen de geesteswetenschappen	29
Conclusies	30
3. <i>PEER REVIEW</i> EN KWALITEITSINDICATOREN	31
<i>Peer review</i> : acceptatie en beperkingen	31
Reflectie op de bezwaren	33
<i>Peer review</i> als basis voor kwaliteitsindicatoren	34
Conclusies	35

4. **REFLECTIE OP BIBLIOMETRISCHE METHODEN** 36
  - Bibliometrie, impactfactoren, Hirsch-index 37
  - De waarde van bibliometrische indicatoren 38
  - Classificaties van tijdschriften en wetenschappelijke uitgeverijen 41
  - Nationale classificatiesystemen 43
  - Conclusies 45
  
5. **NAAR EEN SYSTEEM VAN KWALITEITSINDICATOREN VOOR DE GEESTESWETENSCHAPPEN** 47
  - Randvoorwaarden 47
  - Schets van een systeem van kwaliteitsindicatoren voor de geesteswetenschappen 48
  - Contexten van toepassing 51
    - (a) Onderzoeksvisitatie in het kader van het SEP-protocol 51
    - (b) Beoordelingen van onderzoeksvorstellen en individuele cv's 52
    - (c) Prestatiebeoordelingen binnen faculteiten en instituten 52
  - Vervolgacties en aanbevelingen 53

## **BRONNEN** 55

## **BIJLAGEN**

1. Overzicht indicatoren (input voor pilotfase) 57
2. Enkele buitenlandse voorbeelden van ontwikkelingen rond indicatoren voor geestes- en sociale wetenschappen 59
3. Lijst van geïnterviewden 67



# SAMENVATTING

## Voorgeschiedenis

In 2009 stelde de Commissie Nationaal Plan Toekomst Geesteswetenschappen (de Commissie Cohen) in haar eindrapport (*Duurzame Geesteswetenschappen*) dat er in de geesteswetenschappen sprake is van een ‘inadequaat instrumentarium van kwaliteitsobjectivering en –differentiatie’.<sup>1</sup> De commissie kwam met de volgende aanbeveling voor de KNAW: ‘Neem het voortouw in de ontwikkeling van een maximaal eenvoudig, helder en adequaat systeem van kwaliteitsindicatoren in de Nederlandse geesteswetenschappen.’<sup>2</sup>

De KNAW heeft deze aanbeveling overgenomen en heeft in oktober 2009 een Commissie Kwaliteitsindicatoren Geesteswetenschappen ingesteld met als opdracht:

- een internationale inventarisatie te maken van de bestaande methoden van kwaliteitsbeoordeling in de geesteswetenschappen (inventarisatiefase);
- op grond van deze inventarisatie en na een uitgebreide consultatie van het geesteswetenschappelijke veld een systeem van kwaliteitsindicatoren in de Nederlandse geesteswetenschappen te ontwerpen (conceptuele fase);
- door het uitzetten van pilots de werkzaamheid van het systeem empirisch te toetsen (empirische fase) en op grond van deze uitkomsten het systeem zo nodig aan te passen;
- een eindadvies uit te brengen.

De commissie presenteert in dit rapport een interim-rapportage, voorafgaand aan de pilotfase van het project.

---

1 Commissie Nationaal Plan Toekomst Geesteswetenschappen (2008) 41.

2 *Ibid.*, 45.

De commissie heeft moeten constateren dat bij sommige beleidsmakers te hoge verwachtingen bestaan rond een ‘simpel’, liefst maximaal gekwantificeerd, systeem, dat gemakkelijke vergelijkingen mogelijk maakt tussen onderzoeksgroepen en zelfs hele vakgebieden. Anderzijds constateert de commissie ‘in het veld’ nog steeds een behoorlijke aversie tegen de notie van het ‘meten’ van kwaliteit en tegen het toepassen van wat men ziet als managementmodellen uit het bedrijfsleven, en is er een breed gedragen zorg over de toenemende bureaucratische last van de evaluatiecultuur. De commissie ziet de noodzaak van evaluaties en verantwoording als een gegeven, maar heeft gezocht naar een systeem van kwaliteitsindicatoren dat maximaal adequaat is, in de zin dat het recht tracht te doen aan de veelvormigheid en veelzijdigheid van de wetenschapspraktijk in de geesteswetenschappen en daarbuiten, maar dat ook in de praktijk werkbaar is en niet leidt tot overmatige bureaucratische ballast.

De commissie heeft haar bevindingen gebaseerd op een inventarisatie van internationaal onderzoek over kwaliteitsbeoordelingen en kwaliteitsindicatoren, een inventarisatie van een aantal recente ontwikkelingen in binnen- en buitenland, en interviews met een groot aantal betrokkenen in de Nederlandse context.

## Kwaliteit en kwaliteitsindicatoren

De kwaliteit van wetenschappelijk onderzoek is in meerdere opzichten een relatief of relationeel begrip: zij wordt toegekend door anderen, in relatie tot een bepaalde gebruikscontext en aan de hand van bepaalde externe standaarden. Afhankelijk van de gebruikscontext kan men verschillende aspecten van kwaliteit onderscheiden. Wetenschappelijke kwaliteit geeft de bruikbaarheid van de resultaten van onderzoek aan voor de wetenschappelijke gemeenschap binnen een of meerdere disciplines. Maatschappelijke kwaliteit of maatschappelijke relevantie geeft de bruikbaarheid aan van resultaten van onderzoek voor een bredere groep: de overheid, het bedrijfsleven, culturele en maatschappelijke instellingen. Dit laatste aspect heeft met de toenemende belangstelling voor kennisvalorisatie de laatste jaren aan gewicht gewonnen. Door maatschappelijke kwaliteit nadrukkelijk een plaats te geven in haar rapportage sluit de commissie aan bij het rapport *Impact Assessment Geesteswetenschappen* van NWO (2009)<sup>3</sup> en de *Handreiking Evaluatie van Maatschappelijke Relevantie van Wetenschappelijk Onderzoek* van het ERiC-project van KNAW, NWO, VSNU en HBO-raad (2010).<sup>4</sup>

In de praktijk wordt wetenschappelijke kwaliteit toegekend door *peers*. *Peer review* wordt breed geaccepteerd en ook de commissie is ervan overtuigd dat het een onmisbaar instrument is. *Peer review* kan echter worden ondersteund en versterkt door externe kwaliteitsindicatoren. Dat zijn alle indicaties voor kwaliteit die gevonden kunnen worden buiten de inhoud van de onderzoeksresultaten zelf zoals beoordeeld door *peers*. Externe kwaliteitsindicatoren omvatten zowel output-indicatoren (bijvoorbeeld het publiceren in A-tijdschriften, eventueel ook bepaalde bibliometrische indicatoren)

---

3 Broek & Nijssen (2009)

4 ERiC (2010)

als *esteem*-indicatoren (prijzen, wetenschappelijke functies en andere blijken van externe waardering). Externe kwaliteitsindicatoren ondersteunen het proces van *peer review* door het van een intersubjectieve basis te voorzien en door het te structureren. Een stelsel van kwaliteitsindicatoren is werkbaar, mits het maximaal eenvoudig en economisch is ingericht op het leveren van de benodigde informatie door de te beoordelen partijen. Aan dit aspect, mede in relatie tot het huidige systeem van verslaglegging in METIS, wordt in de pilotfase nadrukkelijk aandacht besteed.

## Kwaliteitsindicatoren in de geesteswetenschappen

Er is geen reden om het begrip onderzoekskwaliteit in de geesteswetenschappen anders te benaderen dan in andere wetenschapsgebieden. Ook in de geesteswetenschappen bestaat kwaliteit in de betekenis van resultaten van onderzoek voor de wetenschappelijke gemeenschap en daarbuiten. En ook hier zijn het in de eerste plaats *peers* die in staat geacht moeten worden om, mede met behulp van externe indicatoren, een kwaliteitsoordeel te vellen. De geesteswetenschappen vragen dan ook niet om een eigen type kwaliteitsindicatoren.

Wel vragen de geesteswetenschappen om een vrij breed scala van indicatoren voor kwaliteit dat recht doet aan de diversiteit aan producten, doelgroepen, en publicatieculturen die men binnen het gebied aantreft. Monografieën en internationale publicaties in andere talen dan het Engels moeten op een faire wijze meegewogen kunnen worden, hetgeen bijvoorbeeld betekent dat het primaat niet kan liggen bij bibliometrische indicatoren die vooralsnog gebaseerd zijn op databestanden die voornamelijk uit Engelstalige tijdschriftpublicaties bestaan. Daarnaast moet het systeem ruimte bieden voor andere vormen van output dan wetenschappelijke publicaties, zoals databases, catalogi en tekstedities. Het systeem dient zo breed te zijn als de wetenschappelijke praktijk vereist, maar tegelijk voldoende flexibel om in verschillende contexten maatwerk te kunnen bieden en werkbaar te blijven. In het benadrukken van het belang van een flexibel systeem met de mogelijkheid om contextspecifieke indicatoren op te nemen (gerelateerd aan de eigenheid van een bepaalde discipline of aan de missie van een bepaald instituut) sluit de commissie aan bij de bevindingen van het KNAW-rapport *Judging Research on its Merits* (2005).

### *Peer review*

Het rapport wijdt om twee redenen een aparte beschouwing aan de voor- en nadelen van *peer review* als basis van kwaliteitsbeoordelingen. Ten eerste omdat kwaliteitsindicatoren moeten functioneren binnen een systeem van *peer review* en bedoeld zijn om de risico's van *peer review* (met name de subjectiviteit) te minimaliseren. Ten tweede omdat veel externe kwaliteitsindicatoren zelf hun basis vinden in *peer review*. Zij zijn als het ware de neerslag van herhaalde beoordelingen door vakgenoten. Kritiek op *peer review* als zodanig kan dus in beginsel ook de waarde van externe indicatoren ondermijnen.

De commissie concludeert dat de meest zwaarwegende bezwaren tegen *peer review* kunnen worden geneutraliseerd door procedurele waarborgen en door het gebruik van externe indicatoren die een intersubjectieve basis geven aan het oordeel van *peers*. De commissie concludeert verder dat de gangbare bezwaren tegen *peer review* slechts in geringe mate afbreuk doen aan de bruikbaarheid van op *peer review* stoelende externe kwaliteitsindicatoren, omdat deze indicatoren veelal gebaseerd zijn op geaggregeerde oordelen van *peers* en omdat zij idealiter niet geïsoleerd, maar in samenhang met andere indicatoren worden gebruikt voor de beoordeling van kwaliteit.

## Bibliometrische indicatoren

Het rapport wijdt eveneens een aparte beschouwing aan het gebruik van bibliometrische indicatoren. Tegen het ongereflacteerd gebruik van bibliometrische indicatoren in het algemeen en in de geesteswetenschappen in het bijzonder is een aantal bezwaren in te brengen die in aard variëren van principieel tot praktisch. Meer principieel: er is geen een-op-een-relatie tussen impact en kwaliteit; er zijn grote verschillen in citatiecultuur tussen vakgebieden, naar onderwerp binnen vakgebieden, en ook tussen taalgebieden (Angelsaksisch of Frans) die vergelijkingen van citatiescores bemoeilijken; en er is een grote *mismatch* tussen de vaak lange ‘halfwaardetijd’ van publicaties in de meeste geesteswetenschappen en het om praktische reden beperkte tijdvenster (van twee tot drie jaar) dat doorgaans bij het berekenen van bibliometrische citatiescores wordt gebruikt.

Meer praktisch: de bestaande databases bieden een volstrekt onvoldoende dekking voor geesteswetenschappelijke publicaties. Op dit laatste punt is recent enige verbetering te zien: ook monografieën en niet-Engelstalige publicaties worden geleidelijk aan in sommige bestanden opgenomen. Niettemin blijven de principiële bezwaren vooralsnog van kracht. Dit sluit niet uit dat in delen van de geesteswetenschappen bibliometrische indicatoren een zekere informatieve waarde kunnen hebben en gebruikt kunnen worden als proxy-variabelen. Maar bibliometrische informatie dient, voor zover überhaupt relevant, in samenhang met andere indicatoren beschouwd te worden, en weging van alle indicatoren door *peers* is essentieel. Er kan geen sprake zijn van automatische calculaties.

Een gedeeltelijk alternatief voor elektronische bibliometrische *rankings* is het classificeren van output naar de status van de publicatiemedia (A-, B- en C-tijdschriften en boekseries). In navolging van initiatieven in onder meer Noorwegen en Vlaanderen ondersteunt de commissie een dergelijke ontwikkeling die ze ziet als een poging om reliëf aan te brengen in de onderzoeksoutput van individuen en instituten. Waar een initiatief om op Europees niveau tot een classificatie van geesteswetenschappelijke tijdschriften te komen is mislukt (ERIH), lijkt het geboden om alsnog een landelijk gedragen classificatie (met internationale *benchmarking*) van nationale en internationale media te maken. De commissie ziet hier vooral een rol voor de nationale onderzoekscholen onder auspiciën en regie van de KNAW.

## Stysteem van kwaliteitsindicatoren

Het door de commissie voorgestelde systeem van kwaliteitsindicatoren (zie hoofdstuk 5 voor een schematische weergave) gaat uit van een beoordeling door *peers* van zowel wetenschappelijke als maatschappelijke kwaliteit (op een schaal van 1-5), aan de hand van telkens drie criteria (output, gebruik van output, indicaties van erkenning). Voor elk van deze criteria wordt vervolgens een aantal indicatoren geformuleerd. Uit deze lijstjes kunnen op de context toegesneden selecties gemaakt worden en zij kunnen desgewenst worden aangevuld (in de categorie 'overige') met disciplinespecifieke of contextspecifieke (bijvoorbeeld missiegerelateerde) indicatoren. Er is ook ruimte voor het toevoegen van nieuwe indicatoren die eventueel voortvloeien uit een toenemende digitalisering van de wetenschaps- en publicatiepraktijk. Het systeem als geheel is daarmee flexibel en maximaal eenvoudig. De selectie van relevante indicatoren wordt overgelaten aan de te beoordelen discipline (bijvoorbeeld bij landelijke visitaties) of het te beoordelen instituut.

Een verdere specificatie van de manier waarop de te beoordelen instanties per indicator de relevante informatie moeten leveren, en een onderzoek naar hoe dit zich verhoudt tot de op dit moment gangbare METIS-registratie van onderzoeksgegevens, moet in de pilotfase plaatsvinden.

## Pilotfase en voorlopige aanbevelingen

Het hier in concept ontwikkelde systeem van kwaliteitsindicatoren zal worden getest in een pilotfase die om te beginnen twee proefevaluaties omvat: bij het Meertens Instituut van de KNAW en bij het Instituut voor Cultuurwetenschappelijk Onderzoek Groningen (ICOG) van de Rijksuniversiteit Groningen. Onder meer zal daarbij gekeken worden naar de differentiërende kracht van het systeem en naar verschillende aspecten van praktische uitvoerbaarheid (zoals de mogelijkheden van *extended peer review*, de mogelijkheden van bibliografische classificaties en de meest economische procedures voor het leveren van indicatoren).

Daarnaast is de commissie in de pilotfase betrokken bij een project van het CWTS dat – in opdracht van het CvB van de EUR – de mogelijkheden in kaart brengt om met verbeterde middelen bibliometrisch onderzoek te doen binnen de discipline van de geschiedenis in Nederland. De commissie stelt zich neutraal op tegenover de bestuurlijke en beleidsmatige doelstellingen van dit project, maar stelt de vraag centraal wat de mogelijkheden zijn van de gekozen nieuwe benaderingen en hoe de balans ligt tussen extra investeringen en meeropbrengst.

Onder voorbehoud van de resultaten van deze pilotfase is de commissie voornemens in haar eindrapport de volgende aanbevelingen te doen aan de KNAW:

- neem het voortouw en de regie voor het initiatief om via de landelijke onderzoekscholen te komen tot een breed gedragen classificatie van de nationale en

internationale publicatiemedia (tijdschriften, boekenreeksen, uitgeverijen) in de geesteswetenschappen;

- agendeer een discussie over kwaliteitsindicatoren in de Adviesraden van de KNAW, om eventuele verschillende opvattingen over kwaliteitsindicatoren in kaart te brengen en de mogelijkheden van verdere consensusvorming te onderzoeken. Dit laatste lijkt met name opportuun door de verrassende convergentie tussen de bevindingen van dit interimrapport en het onlangs verschenen *Rapport Kwaliteitsbeoordeling in de Ontwerpende en Construerende Disciplines*.

# TEN GELEIDE

Voor u ligt een interim-rapport van de Commissie Kwaliteitsindicatoren Geesteswetenschappen van de KNAW. In dit 'Ten geleide' treft u enkele preliminaire opmerkingen aan over de samenstelling en taakopvatting van de commissie en over de opbouw van het rapport.

De Commissie Kwaliteitsindicatoren Geesteswetenschappen is in het najaar van 2009 ingesteld door het bestuur van de KNAW. Op persoonlijke titel werden als leden van de commissie benoemd:

- Prof. dr. Keimpe Algra, voorzitter, hoogleraar Geschiedenis van de Filosofie van de Oudheid en de Middeleeuwen, UU;
- Prof. dr. Hans Bennis, directeur Meertens Instituut KNAW, hoogleraar Taalvariatie, UvA;
- Prof. dr. Jan Willem van Henten, hoogleraar Uitlegging van het Nieuwe Testament, Vroeg-Christelijke Letterkunde en Hellenistisch-Joodse Literatuur, UvA;
- Prof. dr. Mary Kemperink, hoogleraar Moderne Nederlandse Letterkunde, RUG;
- Prof. dr. Wijnand Mijnhardt, hoogleraar Geschiedenis na de Middeleeuwen, i.h.b. Cultuur- Mentaliteits- en Ideeëngeschiedenis, UU;
- Prof. dr. Paul Rutten, hoogleraar Digitale Mediastudies UL (tot 31 december 2010).

De commissie werd ondersteund door een projectteam, bestaande uit:

- Dr. Jack Spaapen, secretaris Raad voor Geesteswetenschappen;
- Dr. Koen Hilberdink, hoofd afdeling Genootschap KNAW;
- Dr. Mark Pen, secretaris/onderzoeker

De commissie zag zich geconfronteerd met verschillende standpunten over nut, noodzaak en mogelijkheden van een systeem van kwaliteitsindicatoren. Enerzijds trof zij bij sommige beleidsmakers te hoge verwachtingen aan met betrekking tot een 'simpel',

lieft maximaal gekwantificeerd, systeem, dat gemakkelijke vergelijkingen mogelijk maakt tussen onderzoeksgroepen en zelfs hele vakgebieden. Anderzijds leeft er in het veld van onderzoekers onmiskenbaar nog een behoorlijke aversie tegen de notie van het ‘meten’ van kwaliteit en tegen het toepassen van wat men ziet als management-modellen uit het bedrijfsleven, en is er een breed gedragen zorg over de toenemende bureaucratische ballast als gevolg van de steeds sterker wordende evaluatie- en verantwoordingscultuur.<sup>5</sup> De commissie heeft geprobeerd beide extreme standpunten te vermijden door te zoeken naar een systeem van kwaliteitsindicatoren dat maximaal adequaat is (en dus niet alleen maar eenvoudig), voor zover het recht tracht te doen aan de veelvormigheid en veelzijdigheid van de wetenschapspraktijk in de geesteswetenschappen en daarbuiten, maar dat ook in de praktijk werkbaar is en niet tot overmatige bureaucratische ballast leidt. Of het ontworpen systeem *de facto* aan deze beide randvoorwaarden voldoet, is nadrukkelijk één van de hoofdvragen die in de pilot-fase van het project aan de orde zullen komen.

De commissie is zich ervan bewust dat ‘geesteswetenschappen’ geen zeer helder gedefinieerd domein is en dat het internationaal niet altijd dezelfde reikwijdte heeft: in sommige landen worden bijvoorbeeld de rechtswetenschap en delen van de antropologie tot de geesteswetenschappen gerekend. De commissie neemt in dit opzicht een pragmatisch standpunt in en volgt de opdracht die spreekt van kwaliteitsindicatoren ‘in de Nederlandse geesteswetenschappen’. Zij neemt als uitgangspunt van reflectie wat in de Nederlandse context onder ‘geesteswetenschappen’ wordt verstaan.<sup>6</sup>

De commissie gevoelde van meet af aan de noodzaak om bij het ontwikkelen van een systeem van kwaliteitsindicatoren de context te betrekken waarbinnen kwaliteitsindicatoren functioneren. Het rapport bevat daarom ook beschouwingen over verschillende contexten van onderzoeksevaluatie en over *peer review*. De opbouw van het rapport is als volgt.

De eerste twee hoofdstukken bieden een oriënterende analyse die het kader moet bieden voor het uiteindelijk te ontwikkelen systeem van kwaliteitsindicatoren. In hoofdstuk 1 wordt ingegaan op het begrip wetenschappelijke kwaliteit en op de rol van kwaliteitsindicatoren bij onderzoeksbeoordelingen. In hoofdstuk 2 wordt de onderzoekspraktijk in de geesteswetenschappen omschreven en wordt nagegaan hoe vanuit deze praktijk moet worden aangekeken tegen het begrip kwaliteit en wat dit betekent voor kwaliteitsbeoordelingen en het gebruik van kwaliteitsindicatoren. Hoofdstuk 3 behandelt voor- en nadelen van *peer review* en de manier waarop kwaliteitsindicatoren *peer review* kunnen ondersteunen. Hoofdstuk 4 richt zich op de vraag naar de bruikbaarheid van bibliometrische methoden (en eventuele alternatieven daarvoor) in de geesteswetenschappen. De inzichten uit hoofdstuk 1 tot en met 4

---

5 Zie bijvoorbeeld Head (2011) 58-64.

6 Zie Commissie Nationaal Plan Toekomst Geesteswetenschappen (2008) 11: ‘De grote familie van disciplines die zich onder deze noemer schaart, richt zich op de studie van talen en culturen, de letteren en de kunsten, geschiedenis en archeologie, godsdiensten, ethiek, gender en filosofie, (...) inclusief belangrijke delen van de communicatie- en mediastudies’.



leveren de bouwstenen voor hoofdstuk 5, waarin een systeem van kwaliteitsindicatoren wordt gepresenteerd en wordt aangegeven hoe dit in de praktijk van kwaliteitsbeoordelingen kan functioneren.

De commissie heeft haar bevindingen gebaseerd op een inventarisatie van internationaal onderzoek over kwaliteitsbeoordelingen en kwaliteitsindicatoren, een inventarisatie van een aantal recente ontwikkelingen in binnen- en buitenland, en een uitgebreide consultatie van het geesteswetenschappelijke veld. Voor een overzicht van een aantal recente buitenlandse ontwikkelingen, zie bijlage 2. Voor een lijst van gesprekspartners, zie bijlage 3.



# 1. KWALITEITSBEOORDELING EN KWALITEITSINDICATOREN

Dit hoofdstuk biedt, in combinatie met het volgende, een inventariserende analyse van begrippen, feiten en ontwikkelingen die van belang zijn voor een goed begrip van de problematiek van kwaliteitsbeoordelingen en kwaliteitsindicatoren in de geesteswetenschappen. Wat verstaan we onder wetenschappelijke kwaliteit, en hoe wordt die toegekend? Waarom is er behoefte aan kwaliteitsindicatoren, en wat zijn dat? Wat verstaan we onder geesteswetenschappen, en welke specifieke kenmerken van deze wetenschappen rechtvaardigen een eigen, toegespitst systeem van kwaliteitsindicatoren? Binnen welke verschillende contexten vinden onderzoeksevaluaties plaats en in hoeverre vragen die verschillende contexten om verschillende benaderingen? De analyse die in deze eerste twee hoofdstukken op basis van de *status quo* geboden wordt, levert een aantal voorwaarden op waaraan het systeem van kwaliteitsindicatoren dat in hoofdstuk 5 wordt ontwikkeld zal moeten voldoen. In dit eerste hoofdstuk richten we de aandacht op het begrip wetenschappelijke kwaliteit, op de contexten waarbinnen kwaliteitsbeoordelingen plaatsvinden, en op de rol van kwaliteitsindicatoren.

## **Kwaliteit van wetenschappelijk onderzoek**

De kwaliteit van producten of diensten is in meerdere opzichten relatief of relationeel. Het is iets wat wordt toegekend *door* anderen, het wordt toegekend *in relatie tot* bepaalde doelstellingen (een gebruikscontext) en het wordt toegekend *in vergelijkend perspectief*. Doordat kwaliteit gerelateerd is aan een gebruikscontext – levert het product wat de gebruikers ervan mogen verwachten? – en bepaald wordt in vergelijking tot andere producten die als ijkpunt (*benchmark*) fungeren, is het mogelijk van *intersubjectief beargumenteerbare* kwaliteit te spreken.

Dit alles geldt ook voor de kwaliteit van wetenschappelijk onderzoek. Ook die wordt toegekend door anderen, voornamelijk vakgenoten of *peers*. Zij gaan na of het onderzoek brengt wat gebruikers ervan mogen verwachten. De wetenschappelijke gemeenschap verwacht nieuwe kennis en nieuwe inzichten, perspectieven voor vervolgonderzoek en nieuwe discussies. Andere gebruikers, zoals de industrie, de overheid, culturele en maatschappelijke instellingen hebben weer hun eigen legitieme verwachtingen. *Peers* worden geacht in staat te zijn al deze verwachtingen, of althans de meeste, te articuleren en te beoordelen of specifiek onderzoek hieraan voldoet. Zij doen dit onder meer door de prestaties van nationale en internationale collega's als *benchmark* te gebruiken. Voor zover hun oordelen beargumenteerbaar zijn, ontstijgen ze het niveau van het subjectieve en intuïtieve.

## Kwaliteitsbeoordelingen in verschillende contexten

De min of meer systematische beoordeling van het werk van onderzoekers door *peers* (*peer review*) heeft haar wortels in de achttiende eeuw, toen voor het eerst in speciale tijdschriften voor onderzoekers over hun wetenschappelijke resultaten werd gecommuniceerd.<sup>7</sup> *Peer review* bepaalde wat al dan niet kon worden gepubliceerd. Deze vorm van *peer review*, de beoordeling van individuele publicaties, wordt nu nog steeds toegepast bij de meeste wetenschappelijke tijdschriften en boekseries.

Tegenwoordig omvat kwaliteitsbeoordeling van wetenschappelijk onderzoek in de meeste contexten echter meer dan de beoordeling van individuele publicaties. Bij visitaties worden complete onderzoeksgroepen en instituten beoordeeld. In Nederland gebeurt dat volgens het *Standard Evaluation Protocol* (SEP). In dergelijke beoordelingen gaat het bovendien niet alleen meer om de geaggregeerde kwaliteit van individuele publicaties, maar spelen bijvoorbeeld ook de richting en de sturing van het onderzoek en de impact van het onderzoek op de wetenschap of de samenleving een rol.<sup>8</sup> Het SEP voorziet verder in een interne *midterm review* halverwege de zesjarige visitatiecyclus, met dezelfde criteria als bij een officiële visitatie, en sommige universiteiten experimenteren daarnaast nog met een vorm van jaarlijkse monitoring via een 'kwaliteitskaart' voor programma's, instituten of faculteiten, eveneens met een breed scala van criteria die ontleend zijn aan het SEP-protocol.

Kwaliteitsbeoordelingen van een andere aard en op een kleinere schaal treffen we aan waar NWO en de EU onderzoeksvoorstellen beoordelen in verband met de toewijzing van onderzoeksubsidies. Op hun verzoek beoordelen referenten zowel – prospectief – zaken als originaliteit van de onderzoeksvragen, relevantie, voorgestelde

---

7 De Solla Price (1963).

8 Het SEP-protocol voorziet in vier beoordelingscriteria: 'quality', 'productivity', 'relevance', en 'vitality and feasibility'. Merk op dat het in dit rapport ontwikkelde systeem van kwaliteits-indicatoren alleen globaal de eerste drie van deze criteria dekt. 'Vitality and feasibility' hebben vooral te maken met de strategische planning en organisatie van het onderzoek, betreffen dus niet direct de kwaliteit van het onderzoek als zodanig en vallen daarmee buiten het bestek van dit rapport. Zie verder de paragraaf 'Contexten van toepassing' in hoofdstuk 5.

methode en haalbaarheid, en – retrospectief – het onderzoeks-cv van de aanvrager.

Ten slotte worden in praktisch alle faculteiten en instituten onderzoeksgroepen en onderzoekers jaarlijks beoordeeld op wetenschappelijke kwaliteit en productiviteit. Meestal gaat het daarbij om een globale beoordeling van de activiteiten van individuele onderzoekers, vaak, maar niet uitsluitend, in R&O-gesprekken in het kader van het personeelsbeleid. Centraal staat daarbij de beoordeling van de onderzoeksoutput in relatie tot de beschikbare onderzoekstijd, die consequenties kan hebben voor de allocatie of herallocatie van onderzoekscapaciteit en dus voor het takenpakket van de individuele medewerker. De criteria bij dit soort individuele beoordelingen zijn doorgaans een afgeleide van de criteria van het SEP-protocol. Er wordt als het ware nagegaan hoe een visitatiecommissie zou aankijken tegen de kwalitatieve en kwantitatieve bijdrage van de individuele onderzoeker aan zijn of haar programma of instituut.

## Standaardisering van *peer review*-procedures

In de meeste van deze contexten vormt *peer review* de basis voor de beoordelingen. In hoofdstuk 3 wordt een verdere analyse gegeven van voor- en nadelen van dit systeem. Voor het moment volstaat het vast te stellen dat *peer review* brede steun geniet als basis voor kwaliteitsbeoordelingen, maar dat het systeem ook op een aantal punten onder druk staat. Om te beginnen heeft de enorme groei van het wetenschappelijk bedrijf, in combinatie met de relatieve krapte van de beschikbare middelen, geleid tot een aanzienlijke toename van het aantal evaluatiemomenten en evaluatiecontexten – hierboven noemden we de drie belangrijkste. Daarnaast leidde een toenemende publicatiedruk tot een forse stijging van het aantal aangeboden manuscripten bij tijdschriften, en in een groeiend aantal tijdschriften, al dan niet in digitale vorm. De aan dit alles verbonden toenemende vraag naar *reviewers*, die op hun beurt zelf ook onder druk staan om te publiceren, zorgt voor een spanning in het systeem van *peer review*, met het risico van kwaliteitsverlies. Afgezien van deze contingente factoren moet *peer review* altijd strijden tegen de schijn van subjectiviteit en van het leunen op intuïtie ('I know quality when I see it'). Mede om deze redenen is de behoefte ontstaan aan gestandaardiseerde procedures. Die vereenvoudigen immers het proces van *peer review* op een bepaalde manier, terwijl ze tegelijkertijd de vergelijkbaarheid van de oordelen vergroten en de subjectiviteit indammen. Om de vergelijkbaarheid en objectiviteit van *peer review*-beoordelingen zo veel mogelijk te garanderen en de procedures te stroomlijnen zijn er dan ook in vele landen initiatieven genomen om criteria en methoden van onderzoeksevaluaties in één systeem onder te brengen.<sup>9</sup>

---

9 In Nederland bestaat sinds 2003 het SEP, in Engeland al langer de Research Assessment Exercise (sinds kort het Research Excellence Framework), in Australië het Research Quality Framework (sinds kort Excellence Research Australia, ERA) en in Frankrijk het Comité National d'Evaluation de Recherche (CNER). Al deze stelsels beogen kwaliteitsbeoordeling te systematiseren (rechtvaardiger te maken, en minder belastend) en effectiever te maken (doordat de resultaten op enigerlei wijze in nationaal perspectief kunnen worden geplaatst).

## Kwaliteitsindicatoren

Op een dergelijke systematisering en standaardisering van de procedures kan een systeem van kwaliteitsindicatoren een belangrijke aanvulling vormen, als het handvatten biedt om kwaliteitsoordelen te onderbouwen. Wat kan men in dit verband verstaan onder kwaliteitsindicatoren? In de context van evaluaties voor wetenschappelijk onderzoek bedoelt men met de term kwaliteitsindicator *alle feitelijk vaststelbare aanwijzingen voor wetenschappelijke en maatschappelijke kwaliteit anders dan de inhoud van de wetenschappelijke output zelf*. In die zin is het oordeel van *peers* zelf ook een kwaliteitsindicator. Waar het gaat om het ontwikkelen van een systeem van kwaliteitsindicatoren heeft de commissie er echter voor gekozen zich te richten op wat men *externe* kwaliteitsindicatoren kan noemen: indicatoren buiten het *peer review*-proces die een controleerbare basis aan dat proces verlenen. Sommige van deze indicatoren, de zogenoemde output-indicatoren, zeggen iets over het niveau van de producten van het onderzoek. Voorbeelden zijn: acceptatie in A-tijdschriften, recensies, citaties en anderszins aantoonbare impact. Andere indicatoren zeggen eerder iets over de status die een onderzoeker of onderzoeksgroep geniet onder *peers*: uitnodigingen om als *keynote speaker* op te treden bij grote congressen, redactielidmaatschappen, lidmaatschappen van prestigieuze commissies, prijzen en onderscheidingen. Deze tweede categorie indicatoren noemt men *esteem indicators*. Een systeem van kwaliteitsindicatoren zal beide typen moeten omvatten.

Deze kwaliteitsindicatoren, die uiteindelijk *peer review* kunnen en moeten ondersteunen, zijn zelf in belangrijke mate gebaseerd op eerdere *peer review*, en dus op kwalitatieve oordelen. Aan citatiescores liggen beslissingen van vakgenoten om een artikel te citeren ten grondslag, opname in een A-tijdschrift reflecteert een positief oordeel van redactie en *referees* en ook *esteem indicators* geven het oordeel van anderen weer. Voor zover kwaliteitsindicatoren *peer review* onderbouwen doen ze dat dus door aan het oordeel een intersubjectief aspect te geven ('ook anderen vinden dat dit goed onderzoek is').

## Recente ontwikkelingen

Drie recente ontwikkelingen in de praktijk van de kwaliteitsbeoordeling van wetenschappelijk onderzoek dienen bij reflectie over een te ontwerpen systeem van kwaliteitsindicatoren betrokken te worden: de opmars van de bibliometrie, de groeiende aandacht voor maatschappelijke relevantie en de diversificatie van beoordelingscontexten.

Om niet alleen een intersubjectieve maar ook een kwantitatieve onderbouwing te kunnen geven aan de kwaliteitsoordelen van *peer review* is men in de laatste decennia in toenemende mate gebruik gaan maken van de in de jaren zeventig van de twintigste eeuw ontwikkelde bibliometrie. Oorspronkelijk ontwikkeld om inzicht te krijgen in wetenschapssociologische processen, hebben bibliometrische instrumenten

geleidelijk aan een andere rol gekregen. Ze worden meer en meer gebruikt als indicatoren die ons in staat moeten stellen de publicatieoutput van wetenschappers en de relatieve kwaliteit daarvan op een kwantitatieve schaal zichtbaar te maken.<sup>10</sup> Voorbeelden zijn: impactfactoren die aan tijdschriften worden toegekend, citatietellingen, diverse citatie-indices. Bibliometristen zijn zelf doorgaans aanzienlijk voorzichtiger in hun claims over de bruikbaarheid en zeggingskracht van deze instrumenten dan sommige beleidsmakers bij overheden, universiteiten en nationale en internationale instellingen voor onderzoeksfinanciering. Beleidsmakers hebben nogal eens een gesimplificeerd beeld van bibliometrie als een eenvoudige methode die door middel van een druk op de knop objectieve *rankings* van onderzoekers en onderzoeksgroepen genereert. Anderzijds treft men bij sommige onderzoekers een even weinig doordachte afwijzing aan van alles wat met ‘meten’ van kwaliteit te maken heeft. Een afgewogen oordeel is dus van belang. Daarom gaat hoofdstuk 4 van dit rapport vrij uitvoerig in op de bruikbaarheid van bibliometrische methoden in het algemeen en in de geesteswetenschappen in het bijzonder.

Een tweede recente ontwikkeling: sinds de laatste decennia van de vorige eeuw is men de maatschappelijke relevantie van wetenschap steeds belangrijker gaan vinden. Deze speelt dan ook nationaal en internationaal een rol in veel kwaliteitsevaluaties en maakt deel uit van het kader dat het SEP biedt aan evaluaties in Nederland. Ook in de geesteswetenschappen is deze tendens waarneembaar. Er is een toenemend besef bij zowel instellingen als overheid dat de geesteswetenschappen een belangrijke bijdrage leveren aan de ontwikkeling van wat de Franse socioloog Pierre Bourdieu ‘cultureel kapitaal’ heeft genoemd (kennis, vaardigheden, vorming).<sup>11</sup> Maar men kan ook denken aan de bijdrage van de geesteswetenschappen aan maatschappelijke discussies, bijvoorbeeld over geschiedenis en identiteit, en aan beleidsvorming op het gebied van cultuur en media. De beoordeling van maatschappelijke kwaliteit van wetenschap is methodologisch complexer dan die van de wetenschappelijke kwaliteit. De data zijn minder robuust, het is soms ingewikkelder ze te verzamelen en er een systeem voor te

---

10 De database van het Institute for Scientific Information (ISI) waarop de eerste bibliometrische onderzoeken gebaseerd waren, bevatte alleen publicaties op het gebied van de natuurwetenschappen (*sciences*). Met de geesteswetenschappen werd in de praktijk van de bibliometrie aanvankelijk geen rekening gehouden. Dit is de laatste tijd enigszins aan het veranderen, maar dat is een langzaam proces. Hoofdstuk 4 gaat uitvoeriger in op de relatie tussen bibliometrie en geesteswetenschappen.

11 Zie AWT (2007) 5-6: ‘Kennis uit de alfa- en gammawetenschappen vindt gretig aftrek in bijvoorbeeld het recht, het onderwijs, de geestelijke gezondheidszorg, de politiek en het beleid. Ook het grote publiek toont veel interesse. Dat blijkt bijvoorbeeld uit de verkoopcijfers van managementboeken en historische studies, of uit de bezoekersaantallen van musea of cultureel erfgoed. Alfa- en gammaonderzoek is ook onmisbaar voor beleidsontwikkeling. Vraagstukken rond cohesie en integratie, intergenerationale solidariteit, internationale betrekkingen, marktwerking, globalisering of onderwijshervormingen zijn niet te behandelen zonder alfa- en gammakennis. Zelfs technische problemen, bijvoorbeeld rond mobiliteit, veiligheid of klimaatverandering, hebben de inbreng van deze vakken hard nodig. Deze problemen kunnen immers niet worden opgelost zonder dat burgers hun gedrag veranderen’.

ontwikkelen, en het is niet vanzelfsprekend dat de relevante informatie door wetenschappelijke vakgenoten altijd op waarde geschat kan worden;<sup>12</sup> bovendien zijn er vele verschillende contexten. Niettemin noopt de gedachte dat kwaliteit mede wordt bepaald door de gebruikscontext er toe om maatschappelijke kwaliteit in onderzoeksbeoordelingen mee te wegen aan de hand van aparte indicatoren.<sup>13</sup>

Een derde relevante ontwikkeling is eerder in dit hoofdstuk al aangestipt, toen gewezen werd op de grote hoeveelheid onderzoeksbeoordelingen die wij vandaag de dag kennen, in verschillende contexten en met verschillende doelstellingen. Universiteiten en organisaties als NWO zijn op zoek naar gegevens aan de hand waarvan ze verantwoorde beslissingen kunnen nemen over prioritering en allocatie van middelen. Onderzoekers en hun beoordelaars willen weten hoe ze het doen ten opzichte van de (internationale) concurrentie. Lokale bestuurders willen weten of de missie van het instituut wordt waargemaakt en of sommige groepen betere resultaten behalen dan andere. Een werkbaar systeem van kwaliteitsindicatoren zal in deze verschillende contexten bruikbaar moeten zijn. Niet alle indicatoren zijn in alle contexten (even) relevant. We zoeken dus naar een goed gevulde 'gereedschapskist' met instrumenten waaruit selecties gemaakt kunnen worden voor verschillende doeleinden en contexten.<sup>14</sup>

Nog een vierde recente ontwikkeling verdient hier aandacht. De digitalisering van het primaire wetenschappelijk proces en de wijze waarop wetenschappelijke kennis gedeeld wordt tussen *peers* en met de samenleving, hebben de potentie om het systeem van wetenschappelijke productie diepgaand te veranderen, met alle mogelijke gevolgen van dien voor de noties van kwaliteit en kwaliteitsindicatoren.<sup>15</sup> Digitalisering maakt processen transparanter, ook dat van wetenschapsbeoefening. Noties van kwaliteit kunnen daarmee meer dan voorheen ook op deze processen zelf van toepassing worden verklaard en niet alleen op eindproducten.<sup>16</sup> Daarnaast genereert digitalisering manieren om ook tijdens het proces van wetenschappelijk onderzoek kennis te delen, bijvoorbeeld door wetenschapsblogs die als onderdelen van de wetenschappelijke output waardering verdienen en kwaliteit kunnen indiceren. Een exclusieve focus op gefinaliseerde artikelen in *peer reviewed* tijdschriften, en in het algemeen op een

---

12 Voor de beoordeling van maatschappelijke kwaliteit van onderzoek kan het daarom zinnig zijn een beroep te doen op zogenoemde *extended peer review*; zie hierover verder hoofdstuk 3.

13 Door maatschappelijke kwaliteit nadrukkelijk een plaats te geven in haar rapportage sluit de commissie aan bij het rapport *Impact Assessment Geesteswetenschappen* van NWO (2009) en de *Handreiking Evaluatie van Maatschappelijke Relevantie van Wetenschappelijk Onderzoek* van het EriC-project van KNAW, NWO, VSNU en HBO-raad (2010).

14 In het benadrukken van het belang van een flexibel systeem met de mogelijkheid om contextspecifieke indicatoren op te nemen (gerelateerd aan de eigenheid van een bepaalde discipline of aan de missie van een bepaald instituut) sluit de commissie aan bij de bevindingen van het KNAW-rapport *Judging Research on its Merits* (2005).

15 Zie onder meer Borgman (2007).

16 Halffabricaten, bijvoorbeeld een voorlopig databestand, kunnen daarmee ook aan kwaliteitsbeoordeling onderworpen worden. Zie Verhaar et al. (2010).



gefixeerde tekst als summum van wetenschapsproductie, is met dit alles niet langer noodzakelijk.<sup>17</sup> Momenteel zijn de implicaties van de digitalisering voor de wetenschappelijke praktijk nog te onduidelijk voor een concrete vertaling in een systeem van kwaliteitsindicatoren, maar het te ontwikkelen systeem dient ruimte te bieden voor toevoegingen en bijstellingen op basis van de genoemde nieuwe ontwikkelingen.

De eerste drie van de hier gesignaleerde ontwikkelingen laten zien dat er een spanning bestaat tussen eenvoud en adequaatheid.<sup>18</sup> De bibliometrie heeft zich gaandeweg verfijnd en is daarmee gecompliceerder geworden. Het meewegen van maatschappelijke relevantie is van belang, maar niet eenvoudig. Een systeem van kwaliteitsindicatoren dat in verschillende evaluatieve contexten bruikbaar is moet, behalve breed en fijnmazig, ook flexibel zijn. Kortom, een goed systeem van kwaliteitsindicatoren kan niet eenvoudig zijn, en een eenvoudig systeem zal slechts grove indicaties geven. Men zal dus moeten streven naar een optimale balans.

## Conclusies

De kwaliteit van de producten van wetenschappelijk onderzoek kan in beginsel bepaald worden door *peers*. Kwaliteitsbeoordelingen door *peers* kunnen plaatsvinden in verschillende contexten en met verschillende oogmerken, zowel prospectief als retrospectief. Retrospectieve kwaliteitsbeoordelingen kunnen worden onderbouwd door gebruik te maken van kwaliteitsindicatoren. Deze onderbouwen het proces door *peer review* van een bredere en intersubjectieve basis te voorzien. Tegelijkertijd maken zij het proces transparanter en daarmee op een bepaalde manier eenvoudiger: niet door de introductie van simpele calculaties op basis van één of enkele indicatoren, maar door structuur en vergelijkbaarheid aan te brengen.

### RANDVOORWAARDEN

Op het systeem van *peer review* en op de bruikbaarheid van de bibliometrie wordt later nog uitvoeriger ingegaan (hoofdstukken 3 en 4). Voor het overige kunnen op basis van de bevindingen van dit eerste hoofdstuk voor een te ontwikkelen systeem van kwaliteitsindicatoren de volgende voorwaarden worden geformuleerd:

---

17 Zo veronderstelt de praktijk van *open review* dat onderzoeksresultaten in eerste instantie digitaal gedeeld worden en openbaar *gereviewd*. Auteurs stellen hun bijdrage vervolgens bij op basis van de kritiek, nader onderzoek door de originele auteurs kan dan tot weer nieuwe versies leiden, en in de loop van de ontwikkeling kan een externe commentator toetreden als auteur. Het productieproces krijgt door dit alles een meer open, dynamisch en gemeenschappelijk karakter.

18 De opdracht van de KNAW aan de commissie spreekt in dit verband simpelweg van het ontwikkelen van 'een systeem van kwaliteitsindicatoren'. Het rapport *Duurzame Geesteswetenschappen* dat aan deze opdracht ten grondslag ligt spreekt echter van een 'maximaal eenvoudig, helder en adequaat systeem'. De commissie heeft dit, gezien de spanning tussen eenvoud en adequaatheid, geïnterpreteerd als een oproep om te komen tot een optimale balans tussen eenvoud en precisie.

- een systeem van kwaliteitsindicatoren dient zowel output-indicatoren te omvatten als *esteem indicators* (indicatoren waarin eerdere oordeelsvorming door *peers* en de waardering van de wetenschappelijke gemeenschap tot uitdrukking komt);
- een systeem van kwaliteitsindicatoren dient breed genoeg te zijn om naast wetenschappelijke kwaliteit ook maatschappelijke relevantie te kunnen meewegen;
- de breedte van een goed systeem van kwaliteitsindicatoren mag niet ten koste gaan van de werkbaarheid;
- het systeem moet door zijn flexibiliteit in afzonderlijke contexten maatwerk mogelijk maken;
- het systeem van kwaliteitsindicatoren dient open te staan voor bijstelling of herziening wanneer de consequenties van de digitalisering voor de wetenschapsbeoefening volledig zijn uitgekristalliseerd.

## 2. ONDERZOEK EN ONDERZOEKSKWALITEIT IN DE GEESTESWETENSCHAPPEN

In dit hoofdstuk wordt een aantal aspecten van de onderzoekspraktijk in de geesteswetenschappen beschreven. Om te beginnen komt daarbij de vraag aan de orde in hoeverre de geesteswetenschappen als gebied een zodanige eigenheid vertonen dat zij vragen om een gebiedsspecifiek systeem van kwaliteitsindicatoren. Vervolgens wordt beargumenteerd waarom de interne diversiteit van de geesteswetenschappen vraagt om een breed en flexibel systeem van indicatoren.

### **De eigenheid van de geesteswetenschappen**

De geesteswetenschappen richten zich op de menselijke cultuur en haar voortbrengselen uit heden en verleden: taal, instituties, religie, filosofie, literatuur, beeldende kunst, architectuur, muziek, film en media. Daarbij is er door de inbreng van een ideologisch (*gender-, race-, class-*) en ethisch perspectief en van meer institutionele en sociologische benaderingen sinds enkele decennia een groeiende interdisciplinariteit, ook in de richting van disciplines buiten de geesteswetenschappen. De geesteswetenschappen hebben zich voor een deel onafhankelijk van elkaar ontwikkeld. Ze vertonen onderling een vorm van 'familiegelekenis', maar delen geen gemeenschappelijke 'essentie'. Dit betekent dat het onmogelijk is 'de' geesteswetenschappen in het algemeen te karakteriseren. Niet overal in de wereld rekent men dezelfde disciplines tot de geesteswetenschappen. Ook in de Nederlandse context hebben begrippen als

‘geesteswetenschappen’, *humanities* of ‘*humaniora*’ niet altijd hetzelfde bereik en kan de institutionele positionering van vakken als religiewetenschap, wijsbegeerte en archeologie verschillen.

Het ontbreekt aan één gemeenschappelijk kenmerk waarmee geesteswetenschappen zich van andere wetenschapsgebieden onderscheiden. Veeleer is er een conglomeraat van verschillende vakgebieden met ieder een eigen plaats in het totaal der wetenschappen en met wisselende dwarsverbanden met andere gebieden. Sommige deelgebieden van de geesteswetenschappen, zoals de taalkunde, vertonen in hun werkwijze grote overeenkomsten met de empirische natuurwetenschappen. Er is sprake van vorming en modificatie van theorieën, toetsing van hypothesen aan de empirie en de cumulatieve productie van een gemeenschappelijk gedragen *body of knowledge*. Andere geesteswetenschappen, zoals delen van de geschiedwetenschap, hebben meer verwantschap met bepaalde takken van de sociale wetenschappen. Weer andere richten zich vooral op de interpretatie van individuele en unieke objecten (kunstwerken, literaire of religieuze teksten). Sommige takken van de wijsbegeerte vertonen overeenkomsten met onderdelen van de wiskunde.

Doordat de geesteswetenschappen geen eenheid vormen met een gedeelde identiteit, kan men die veronderstelde identiteit ook niet aanvoeren als reden waarom kwaliteitsbeoordeling voor de geesteswetenschappen als geheel anders zou moeten verlopen dan in andere gebieden. De feitelijke onderzoekspraktijk ondersteunt deze conclusie. Bij alle verschillen in methodologie en onderzoekspraktijk, is men het binnen de geesteswetenschappen in beginsel verregaand eens over wat wetenschappelijk onderzoek beoogt. Dat verschilt niet wezenlijk van de doelstellingen van andere wetenschapsgebieden. Net als in andere wetenschappen gaat het bij wetenschappelijk onderzoek in de geesteswetenschappen om het methodisch vermeerder en verhelderen van onze kennis van de werkelijkheid en van onze relatie tot die werkelijkheid. Net als elders kan onder wat met onderzoek beoogd wordt, ook maatschappelijke relevantie vallen: regelmatig wordt wetenschappelijke kennis gedeeld met een bredere doelgroep van beleidsmakers, cultuurinstellingen, onderwijsprofessionals en overige geïnteresseerden. In al die onderling verschillende geesteswetenschappen wordt bovendien onderzoekskwaliteit in algemene zin op dezelfde wijze opgevat als in andere gebieden. Ook hier bestaat kwaliteit in de betekenis van resultaten van onderzoek voor de wetenschappelijke gemeenschap en daarbuiten. En ook hier zijn het in de eerste plaats *peers* die in staat geacht worden om, mede met behulp van indicatoren, een kwaliteitsoordeel te vellen. Een groot deel van de kwaliteitsindicatoren die elders toepasbaar zijn is daarom ook voor de geesteswetenschappen zonder meer bruikbaar. Alle geesteswetenschappelijke vakgebieden hebben toptijdschriften, belangrijke congressen waar men als *keynote speaker* gevraagd kan worden, onderscheidingen en prijzen. Vanuit dit perspectief bezien zal het te ontwerpen systeem van kwaliteitsindicatoren niet wezenlijk anders zijn dan wat men voor andere wetenschapsgebieden zou kunnen formuleren.

## Diversiteit binnen de geesteswetenschappen

De geesteswetenschappen vragen dus niet om een eigen *type* kwaliteitsindicatoren. Wel vraagt de interne diversiteit van de geesteswetenschappen – een diversiteit die men ook bij andere wetenschapsgebieden aantreft – om een breed *scala* van indicatoren. Drie vormen van diversiteit – wetenschapssociologische verschillen, verschillen in doelstellingen en producten, en verschillen in publicatiekanalen – zouden hun weerslag moeten vinden in een brede waaier van kwaliteitsindicatoren, waaruit vervolgens per discipline, en soms per context, een keuze gemaakt kan worden.

Wetenschapssociologische verschillen kunnen ertoe leiden dat bepaalde indicatoren voor het ene vakgebied relevanter zijn dan voor het andere. In het ene gebied werken bijvoorbeeld vele onderzoekers tegelijkertijd aan de uitbouw en verfijning van dezelfde theorie, terwijl in het andere gebied individuele onderzoekers zich richten op uiteenlopende onderwerpen. Dit maakt een groot verschil voor de toepasbaarheid van bibliometrische indicatoren (zoals citatieanalyses) en stelt grenzen aan de bibliometrische vergelijkbaarheid van (delen van) verschillende disciplines.

Verder kunnen er relevante verschillen zijn in de doelstellingen en producten van het onderzoek. De beoefenaren van een aantal geesteswetenschappelijke disciplines zijn hoeders, ontsluiters en interpretatoren van internationaal en nationaal erfgoed. Net als onderzoek in sommige andere gebieden (bijvoorbeeld de technische wetenschappen) vertaalt geesteswetenschappelijk onderzoek zich niet uitsluitend in de 'klassieke' vorm van artikelen in wetenschappelijke tijdschriften, maar ook in andere producten. Veel onderzoekers zijn actief in de opbouw van databestanden en het ontsluiten van datacollecties – soms voor vakgenoten, soms voor een breder publiek. Zij dragen zo in belangrijke mate bij aan de opbouw van een geesteswetenschappelijke kennisinfrastructuur. Dit betekent dat er bij kwaliteitsbeoordelingen in sommige geesteswetenschappelijke disciplines behoefte is aan een bredere selectie van relevante indicatoren dan louter wetenschappelijke artikelen.

Een derde punt betreft de publicatiekanalen. Voor de meer synthetiserende geesteswetenschappen zijn – naast wetenschappelijke tijdschriften – monografieën en thematische bundels een belangrijk communicatiekanaal. De publicatielijst van een hooggewaardeerde onderzoeker bevat in veel geesteswetenschappelijke disciplines op zijn minst enkele spraakmakende wetenschappelijke boeken. Monografieën zijn dus voor de meeste geesteswetenschappen relevante indicatoren.<sup>19</sup> Waar monografieën relevante kwaliteitsindicatoren zijn, zijn vervolgens bibliometrische indicatoren vaak juist weer minder relevant, omdat die vooralsnog vooral op databases zijn gebaseerd waarin geen monografieën voorkomen. Verder hebben sommige takken van de geesteswetenschappen een zodanig geografisch afgebakend studieobject (de Nederlandse regionale geschiedenis, de Italiaanse of Hongaarse poëzie) dat Engels als publicatietaal niet vanzelfsprekend is. Om op deze gebieden internationaal mee te kunnen doen, is

---

19 Dit kunnen, wellicht in toenemende mate, ook digitale monografieën zijn. Zie Adema (2010).

het publiceren in andere talen dan het Engels vereist. Ook in deze gevallen zijn bibliometrische indicatoren maar beperkt bruikbaar, althans voor zover deze nog exclusief of voornamelijk gebaseerd zijn op citaties van artikelen in Engelstalige tijdschriften.<sup>20</sup>

## Conclusies

In het nastreven van zowel wetenschappelijke kwaliteit als maatschappelijke relevantie zijn de geesteswetenschappen vergelijkbaar met de meeste andere wetenschapsgebieden. Dat geldt ook voor de manier waarop het begrip 'kwaliteit' in de praktijk gehanteerd wordt. In deze opzichten vragen de geesteswetenschappen dus niet om een eigen *type* indicatoren.

Wel zal een adequaat instrumentarium van criteria en indicatoren breed genoeg moeten zijn om recht te doen aan de diversiteit aan benaderingen, doelstellingen, doelgroepen en publicatiekanalen die men binnen de geesteswetenschappen aantreft. Een systeem van indicatoren dat vrijwel alleen onderzoeksoutput meet in de vorm van Engelstalige artikelen in Angelsaksische tijdschriften die bibliometrisch met vrucht onderzocht kunnen worden, is voor de meeste geesteswetenschappelijke disciplines niet adequaat.

De commissie benadrukt nogmaals dat veel andere wetenschapsgebieden een vergelijkbare diversiteit kennen en dat ook daar een beperkt systeem van indicatoren niet adequaat is.

### RANDVOORWAARDE

De bevindingen van dit tweede hoofdstuk leveren voor een te schetsen systeem van kwaliteitsindicatoren de volgende voorwaarde op:

- de geesteswetenschappen vragen om een breed scala van kwaliteitsindicatoren met het oog op de diversiteit in
  - onderzoekspraktijken en publicatieculturen;
  - publicatietalen en publicatiemedia;
  - producten van onderzoek;
  - doelgroepen.

De diversiteit op de eerste twee punten stelt grenzen aan de toepasbaarheid van bibliometrische indicatoren. De laatste twee punten zijn complementair en vragen om een relatief breed scala van kwaliteitsindicatoren.

---

<sup>20</sup> Voor een verdere bespreking van de grenzen aan de bruikbaarheid van bibliometrische indicatoren, zie hoofdstuk 4.

# 3. PEER REVIEW EN KWALITEITSINDICATOREN

Waar in het voorafgaande is gesteld dat kwaliteitsindicatoren *peer review* kunnen ondersteunen en aanvullen, is het nu relevant na te gaan op welke punten *peer review* een dergelijke onderbouwing behoeft en wat de grenzen zijn van de mogelijkheden van kwaliteitsindicatoren in dit verband. Daarnaast is in het voorafgaande geconstateerd dat kwaliteitsindicatoren op vormen van *peer review* gestoeld zijn. Dit betekent dat relevante kritiek op *peer review* ook de waarde van kwaliteitsindicatoren ter discussie kan stellen. Twee redenen dus om in dit rapport een beknopte reflectie op het verschijnsel *peer review* in te lassen. Dit hoofdstuk gaat dan ook na hoe *peer review* wordt toegepast, op welke punten het vatbaar is voor kritiek, hoe aan die kritiek tegemoet gekomen kan worden en wat de consequenties van een en ander zijn voor de rol en de waarde van kwaliteitsindicatoren.

## **Peer review: acceptatie en beperkingen**

Binnen de wetenschapsector bestaat er een groot draagvlak voor *peer review* als methode voor kwaliteitsbeoordeling; *peer review* wordt dan ook ingezet in alle majeure evaluatiecontexten. Idealiter wordt bij *peer review* het oordeel gegeven door experts en is het intersubjectief en beargumenteerd. Door de verschillende manieren waarop *peer review* georganiseerd kan worden, zijn deze voordelen echter niet altijd in gelijke mate aanwezig.

De *peers* zijn soms inderdaad experts, maar in andere gevallen is sprake van brede panels, al dan niet ondersteund door referentoordeelen van specialistische vakgenoten. De procedure voor de selectie van *peers* is niet altijd transparant en niet altijd zijn *peers* wat zij zouden moeten zijn: excellente en ervaren onderzoekers die zelf hoog

scoren op de indicatoren die zij geacht worden toe te passen en die in staat zijn met een scherpe blik *sample*-publicaties en onderzoeksvoorstellen te lezen en indicatoren op waarde te schatten en onderling te wegen.<sup>21</sup> *Peer review* geschiedt soms inderdaad collectief – door een visitatiecommissie, een beoordelingspanel, meerdere *referees* – maar in andere contexten geeft slechts één individu een oordeel: soms geschiedt de beoordeling van een artikel door een enkele *referee*; de voortgang van het onderzoek van een medewerker zal vaak worden beoordeeld door een enkele onderzoeksdirecteur. De onderbouwing van *peer review* is bovendien niet altijd even uitvoerig en expliciet, en soms zelfs afwezig, wat de waarde van *peer review* aantast.

Een zekere bandbreedte op de punten van intersubjectiviteit en vakspecifieke expertise kan in de praktijk aanvaardbaar zijn, afhankelijk van context en doel van de beoordeling en van de mogelijkheid van procedurele compensatie.<sup>22</sup> Ook binnen deze acceptabele bandbreedte kan er echter toch het een en ander misgaan. Verschillende *peers* kunnen op grond van hetzelfde materiaal tot verschillende conclusies komen, op basis van verschillen in achtergrond, kennis, voorkeuren of inzet (*sloppy refereeing*). In de literatuur en in recente discussies treft men dan ook een aantal kritische kanttekeningen aan bij de bruikbaarheid en betrouwbaarheid van *peer review*. De belangrijkste daarvan vatten we hier genummerd samen.<sup>23</sup>

1. *Peer review* kan stoelen op *subjectieve voorkeuren* voor een bepaald type onderzoek, vooral waar de gehanteerde criteria nogal eens impliciet zijn.
2. *Peers* kunnen de neiging hebben de voorkeur te geven aan werk dat *aansluit bij bestaande inzichten en paradigma's*. Dit kan een rem zijn op controversieel onderzoek en wetenschappelijke vernieuwing.
3. *Peer review* kan negatieve consequenties hebben voor *interdisciplinair onderzoek* dat zich op de grenslijnen van verschillende disciplines afspeelt.
4. *Gedeelde belangen en wederzijdse afhankelijkheid* kunnen in een kleine wetenschappelijke gemeenschap een basis zijn voor non-interventie. De *peer* die vandaag een artikel van een onderzoeker beoordeelt, kan morgen zelf door die onderzoeker worden beoordeeld. Het oordeel van de *peer* kan dan 'vals positief' zijn.
5. Of er zijn *tegengestelde belangen*. De *peer* kan onderzoek dat erg lijkt op zijn eigen onderzoek tegenwerken of vertragen, zodat hij als eerste over het resultaat kan publiceren. In dat geval kan het oordeel van de *peer* 'vals negatief' zijn.
6. *Peers* kunnen een voorkeur hebben voor het werk van *gevestigde onderzoekers*,

---

21 Dit heeft ongetwijfeld mede te maken met de toenemende behoefte aan *peer review* die signaleerd werd in hoofdstuk 1.

22 Individuele, en dus niet-intersubjectieve *peer review* door een onderzoeksdirecteur of leidinggevende kan acceptabel zijn omdat deze zich doorgaans tot hoofdlijnen beperkt (functioneert betrokkene naar behoren?) en zich verder kan verlaten op de oordelen van visitatiecommissies of *midterm panels* of andere informatie. Bredere, niet-disciplinespecifieke panels kunnen acceptabel zijn waar zij ondersteund worden door referentrappporten die wel geschreven zijn vanuit vakspecifieke expertise (NWO).

23 Zie bijvoorbeeld Hemlin (2009).



hetgeen leidt tot een stapeling van positieve oordelen en onderzoeksgelden bij onderzoekers met een goede reputatie. Dit heet het Mattheüs-effect ('Want wie heeft zal nog meer krijgen, en wel in overvloed, maar wie niets heeft, hem zal zelfs wat hij heeft nog worden ontnomen.' Mat. 25:29).<sup>24</sup>

7. De procedure voor het selecteren van *peers* is dikwijls niet transparant en kan manipulatieve aspecten hebben.
8. Als *vakgenoten* zijn *peers* niet vanzelfsprekend in staat tot het beoordelen van de *maatschappelijke* kwaliteit en impact van wetenschappelijk onderzoek.
9. Een laatste bezwaar is van meer praktische aard, maar wel zwaarwegend. Door een steeds sterker wordende prestatiecultuur en verantwoordingsdruk in de wetenschapssector moeten vooraanstaande *peers* steeds meer tijd besteden aan de beoordeling van onderzoekers en onderzoeksaanvragen. Dit leidt tot overbelasting van bepaalde *peers*, met het risico dat zij artikelen en voorstellen niet altijd serieus beoordelen. De zorgvuldigheid van het *peer review*-proces kan daardoor onder druk komen te staan.<sup>25</sup>

## Reflectie op de bezwaren

Niet alle genoemde bezwaren spelen in alle contexten een rol en ze zijn bovendien niet van een gelijk gewicht. Bezwaren (4) en (5) betreffen *peer review* van individuele publicaties en slaan op integriteit. Misstappen bij de beoordeling van individuele publicaties komen natuurlijk voor, maar de commissie heeft de indruk dat dit soort *peer review* meestal te goeder trouw plaatsvindt. Bovendien zal *peer review* van individuele publicaties doorgaans zijn beslag krijgen in een leesrapport, hetgeen toch dwingt tot objectivering. In veel gevallen zijn er bovendien meerdere *referees*. Het risico lijkt daarmee aanvaardbaar, en er is geen beter alternatief. Bezwaar (6) betreft *peer review* bij onderzoeksvoorstellen en subsidieaanvragen. Het is echter niet zonder meer duidelijk hoeveel schade het genoemde Mattheüs-effect in een dergelijke context aanricht en of er wel een serieus probleem is. Het komt er immers op neer dat aan een onderzoeker krediet wordt verleend op grond van prestaties geleverd in het verleden, en tot op zekere hoogte is dat precies de bedoeling van dit type onderzoeksbeoordelingen.

De algemene bezwaren (1) t/m (3) lijken het meest serieus. Er staan echter procedurele middelen ter beschikking om aan deze bezwaren tegemoet te komen. Bezwaar (1) kan in de context van beoordelingen van individuele publicaties voor een belangrijk deel worden geneutraliseerd door een leesrapport te vragen; in contexten waarbij een heel cv of een hele groep wordt beoordeeld, kan men verlangen dat de beoordeling onderbouwd wordt door intersubjectieve kwaliteitsindicatoren.

Ook de risico's van (2) en (3) kunnen door aanpassingen in de procedure tot een aanvaardbaar minimum worden teruggebracht. In elk geval daar waar men werkt met bredere panels of met meerdere *referees* kan men bij de selectie van beoordelaars

---

24 Merton (1968) 56-63.

25 Zie bijvoorbeeld KNAW-commissie Kwaliteitszorg (2008).

openheid creëren voor vernieuwing in het wetenschapsdomein en voor interdisciplinariteit. *Peers* met een excellente reputatie op een sterk gespecialiseerd vakgebied, kunnen aangevuld worden met *peers* met een bredere en meer interdisciplinaire kijk. Door panels breder samen te stellen (en enkelvoudige *peer review* te vermijden) kan men ook ten dele tegemoetkomen aan bezwaar (7).

Aan bezwaar (8) kan tegemoet gekomen worden door te werken met wat men *extended peer review* noemt: *peer review*, aangevuld met informatie vanuit een breder domein van deskundigen dan alleen de vakgenoten. Zo kan men maatschappelijke *stakeholders* bij het beoordelingsproces betrekken, hetzij via externe rapportage, hetzij door hen een rol te geven in de beoordelingspanels naast de eigenlijke *peers*.<sup>26</sup>

Ten aanzien van bezwaar (9) merkt de commissie op dat het essentieel is overbelasting van *peers* tegen te gaan. Een afname van beoordelingsmomenten en vereenvoudiging van processen is hoogst gewenst. Men kan proberen doublures te voorkomen en evaluatieresultaten in meerdere contexten te gebruiken (zoals inmiddels gebeurt met de resultaten van de SEP-evaluaties bij de accreditering van onderzoekscholen door de ECOS). Daarnaast kan een overzichtelijk en breed gedragen systeem van kwaliteitsindicatoren het werk zowel van de *peer review committees* als van de te beoordelen onderzoekers en onderzoeksgroepen vereenvoudigen door het te structureren. Voorwaarde is wel dat het systeem flexibel genoeg is voor maatwerk in specifieke contexten, maar de te hanteren kwaliteitsindicatoren en de gebruikte systemen van onderzoeksverslaglegging (zoals METIS) moeten ook voldoende op elkaar aansluiten. Hoofdstuk 5 zal nader ingaan op deze voorwaarden en op de mogelijkheden ze te realiseren.

## Peer review als basis voor kwaliteitsindicatoren

Wegen de bezwaren die er tegen *peer review* kunnen worden ingebracht zodanig zwaar dat ze de waarde en betrouwbaarheid van de verschillende soorten kwaliteitsindicatoren die zelf op *peer review* stoelen (acceptatie in A-tijdschriften, citaties, recensies, verschillende *esteem*-indicatoren) aantasten? De commissie meent van niet. Ten eerste zijn genoemde indicatoren doorgaans gebaseerd op een *aggregatie* van individuele *peer review*-beslissingen, waardoor subjectieve vooroordelen al zoveel mogelijk worden geneutraliseerd. Ten tweede functioneren indicatoren in de praktijk niet individueel maar in clusters met andere indicatoren, zodat ook langs deze weg een voldoende brede basis voor de oordeelsvorming wordt gegarandeerd. Kwaliteitsindicatoren kunnen daarom, hoewel ze uiteindelijk zelf op *peer review* stoelen, op hun beurt gebruikt worden om *peer review*-processen te ondersteunen en te onderbouwen. Zij doen dit, zoals gezegd, door de intersubjectieve basis voor oordeelsvorming te verbreden.

---

<sup>26</sup> Zie ERIC (2010). De mogelijkheid en werkbaarheid van het inschakelen van maatschappelijke *stakeholders* in het proces van *peer review* is een van de aspecten die in de pilotfase van dit project getest zullen worden. Zie daarvoor verder hoofdstuk 5.

## Conclusies

De meeste onderzoekers accepteren *peer review* als de kern van wetenschappelijke kwaliteitsbeoordelingsprocessen en ook de commissie is van mening dat *peer review*, hoewel niet feilloos, het beste instrument is dat we hebben. Bij de toepassing moet wel rekening worden gehouden met bepaalde risico's, die voor een deel kunnen worden teruggebracht door waarborgen in de procedures. Daarnaast kan een systeem van intersubjectieve kwaliteitsindicatoren het proces ondersteunen. Zulke indicatoren vormen als het ware de 'neerslag' van herhaalde *peer review*-oordelen. Daarmee kunnen ze het oordeel van een enkel individu of van een commissie versterken en onderbouwen. Een helder en breed toepasbaar systeem van kwaliteitsindicatoren kan bovendien, mits aan bepaalde voorwaarden van werkbaarheid is voldaan, het werk voor zowel *peer review committees* als voor de te beoordelen groepen vereenvoudigen.

### RANDVOORWAARDE

Hiermee wordt nog een nieuwe randvoorwaarde gegeven waaraan een te ontwerpen systeem van kwaliteitsindicatoren moet voldoen:

- een werkbaar systeem van kwaliteitsindicatoren moet maximaal eenvoudig toepasbaar zijn, zowel voor de instituten die de relevante informatie moeten aandragen als voor de *peers* die de indicatoren moeten wegen.

# 4. REFLECTIE OP BIBLIOMETRISCHE METHODEN

Bibliometrische databases werden voor het eerst opgezet in de jaren zestig van de vorige eeuw. Oorspronkelijk waren ze bedoeld om sneller informatie te kunnen opsporen in de groeiende hoeveelheid literatuur en ook om onderzoek te doen naar communicatiepatronen en samenwerkingsverbanden in de wetenschap. Door te kijken naar citatiepatronen wilde men een beeld krijgen van wetenschapssociologische ontwikkelingen. In de loop van de jaren werden de databases echter ook steeds meer gebruikt voor de beoordeling van de kwaliteit en impact van onderzoek. Momenteel worden citaties in de biomedische en bètawetenschappen min of meer rechtstreeks in verband gebracht met onderzoekskwaliteit. Veel geciteerd worden is daar inmiddels synoniem aan wetenschappelijke kwaliteit. Andere wetenschapsgebieden staan onder een zekere druk om zich eveneens aan beoordelingen met bibliometrische indicatoren te onderwerpen. Zoals eerder aangegeven bestaat het voordeel van bibliometrische indicatoren in de ogen van veel beleidsmakers in hun objectiviteit en veronderstelde eenduidigheid. Bibliometristen oordelen zelf doorgaans genuanceerder over de directe en eenvoudige toepasbaarheid van dit type indicatoren voor onderzoeksbeoordelingen. Zowel op theoretische als op praktische gronden zijn vraagtekens te zetten bij grootschalig en ongerefleeteerd gebruik. In dit hoofdstuk wordt een aantal bibliometrische instrumenten besproken en wordt op een rij gezet wat de bibliometrie momenteel wel en niet kan zeggen over wetenschappelijke kwaliteit in het algemeen en over kwaliteit in de geesteswetenschappen in het bijzonder.

## Bibliometrie, impactfactoren, Hirsch-index

Bibliometristen tellen publicaties en citaties en kunnen zo uitspraken doen over productiviteit en de mate waarin onderzoekers worden geciteerd. Zij baseren zich met hun tellingen op databases met artikelen in een groot aantal wetenschappelijke tijdschriften, zoals het *Web of Science* (WoS) van Thomson Reuters.<sup>27</sup> Verschillende soorten berekeningen genereren verschillende kwantitatieve indicatoren, zoals het aantal citaties per publicatie of relatieve citatiescores in het vakgebied waarbinnen gepubliceerd wordt. Afhankelijk van het gebruikte tijdsvenster (het aantal jaren waarover gemeten wordt) kunnen citatiescores uiteraard variëren.

Een veelgebruikte kwaliteitsindicator is daarnaast de impactfactor van het tijdschrift waarin wordt gepubliceerd. De impactfactor van een tijdschrift laat zien hoe vaak artikelen uit dat tijdschrift gemiddeld zijn geciteerd (veelal berekend over de voorgaande twee jaar). Een impactfactor 3 betekent dat de artikelen in het desbetreffende tijdschrift in de jaren t-1 en t-2 gemiddeld drie keer werden geciteerd. Een impactfactor is een gemiddelde en zegt dus niets over de bandbreedte tussen veel en weinig geciteerde artikelen. Daarnaast is het zo dat tijdschriften die wekelijks verschijnen en die gaan over thema's die wetenschappelijk in de mode zijn (bijvoorbeeld *Nature* en *Science*) een aanzienlijk hogere impactfactor hebben dan tijdschriften die veel minder frequent verschijnen, voor relatief kleine gespecialiseerde vakgebieden.<sup>28</sup>

Ontevreden over de gangbare citatiematen introduceerde de fysicus Jorge E. Hirsch in 2005 een nieuwe en relatief eenvoudige citatiemaat: de Hirsch-index (of H-index). Dit is een vakafhankelijke index die de 'carrière-impact' van een onderzoeker laat zien. Citatiescores van een onderzoeker worden vaak vertekend door enkele veel geciteerde artikelen. De H-index – gebaseerd op dezelfde databases (zoals *Web of Science*) als gewone citatietellingen – gaat uit van de impact van het gehele oeuvre van een onderzoeker. Een wetenschapper heeft index H als H van zijn of haar in totaal N publicaties tenminste H maal geciteerd zijn in andere publicaties, en de andere (N-H) publicaties minder dan H maal geciteerd zijn. Oftewel: een H-index van 10 laat zien dat tien van de publicaties van een onderzoeker tenminste tien keer zijn geciteerd. Volgens Hirsch geeft deze index een realistischer beeld van de wetenschappelijke impact van iemands publicaties dan andere cijfers zoals het totaal aantal citaten (dat beïnvloed kan worden door een klein aantal veelgeciteerde artikelen, zoals overzichtsartikelen/reviews) of het aantal citaten per artikel. De keerzijde is dat een zeer invloedrijke onderzoeker die relatief weinig publiceert maar wel vaak geciteerd wordt, een lage H-index heeft. De H-index is immers een afgeleide van zowel impact als productiviteit. De H-index werkt

---

<sup>27</sup> Het *Web of Science* was voorheen eigendom van het Institute for Scientific Information (ISI) en bekend onder de namen *Science Citation Index*, *Social Science Citation Index* en *Arts and Humanities Index*.

<sup>28</sup> Tijdschriften als *Nature* en *Science* hebben een impactfactor van ongeveer 30, terwijl bijvoorbeeld het belangrijkste tijdschrift voor de natuurkunde (*Physical Review Letters*) een impactfactor heeft van rond de 7.

verder in het nadeel van jonge onderzoekers, en hij is niet consistent: *rankings* blijven niet dezelfde als twee te vergelijken onderzoekers in dezelfde mate extra citaties krijgen (een probleem dat ook bij de *Journal Impact Factor* optreedt).<sup>29</sup>

De H-index is inmiddels misschien de bekendste en meest gebruikte bibliometrische indicator, maar er zijn intussen verschillende andere multidimensionale indices ontworpen, hetzij als complement voor de H-index, hetzij om voor andere gespecificeerde contexten preciezere metingen mogelijk te maken. Zo zijn er nu ook G-, A-, R- en AR-indices.<sup>30</sup> Er is dus een voortdurende uitbreiding en verfijning van het systeem van bibliometrische indicatoren aan de gang.

## De waarde van bibliometrische indicatoren

Bij het gebruik van bibliometrische indicatoren voor het beoordelen van wetenschappelijke kwaliteit is op een aantal punten voorzichtigheid geboden. Het simpelweg en ongewogen tellen van *publicaties* geeft in feite alleen een maat voor productiviteit, niet voor kwaliteit. Het tellen van *citaties* is al weer wat informatiever, maar geeft evenmin een directe indicatie van de kwaliteit, hoogstens van de impact. Zelfs als men corrigeert voor zelfcitaties (en de mogelijkheid van ‘citatieclusters’ van bevriende onderzoekers buiten beschouwing laat) kan men constateren dat verwijzing naar het werk van iemand anders in principe alleen aangeeft dat de desbetreffende onderzoeker dat werk heeft gebruikt. Of de verwijzing positief of negatief is, wordt niet zichtbaar. Nu kan men stellen dat ook een kritische verwijzing aangeeft dat het desbetreffende werk in elk geval belangrijk genoeg gevonden wordt om te worden bestreden en er is zeker wat voor te zeggen om dit soort impact toch mee te wegen bij kwaliteitsbeoordelingen. Er is echter geen een-op-een-relatie tussen hogere citatiescore en hogere kwaliteit.

Dit wordt eens te meer duidelijk als men in ogeschouw neemt dat er belangrijke verschillen zijn in citatiepatronen en citatieculturen tussen verschillende wetenschapsgebieden en zelfs tussen delen van één en hetzelfde wetenschapsgebied. In sommige gebieden werken veel onderzoekers aan dezelfde typen onderzoeksvragen, hetgeen logischerwijze leidt tot een druk ‘citatieverkeer’. Specialistische deelgebieden of onderwerpen waar maar weinig wetenschappers onderzoek naar verrichten genereren veel minder citaties. In het ene vakgebied is het meer *bon ton* om veel te citeren dan in het andere, en er lijkt ook nog een verschil te zijn tussen de Angelsakische praktijk en, bijvoorbeeld, de Franse.<sup>31</sup> Zulke verschillen maken het riskant om in een evaluatieve context citatiescores van onderzoekers en impactfactoren van tijdschriften uit verschillende wetenschapsgebieden te vergelijken. Daarnaast moet men ook nog rekening houden met verschillende scores van verschillende *typen* publicaties. Zo hebben *review articles* een grotere citatie-impact dan gewone artikelen. Ook overzichtsartikelen en handboeken worden vermoedelijk vaker geciteerd dan specialistische stu-

---

29 Waltman & Van Eck (2009)

30 Zie bijvoorbeeld Jin et al. (2007) 855-863; Costas & Bordons (2008) 267-288.

31 Dolan (2007) 28.

dies. In al deze gevallen staat een hogere impact niet gelijk aan een hogere kwaliteit.

Op basis van deze overwegingen moeten we dus sceptisch zijn over het automatische, ongerefl ecteerde gebruik van citatiescores als *directe* en eenduidige indicatie van het kwaliteitsniveau van onderzoekers of groepen.

Een bijkomend probleem bij de toepassing van bibliometrische methoden in de geesteswetenschappen is de ‘halfwaardetijd’ van publicaties die in veel geesteswetenschappen aanzienlijk langer is dan in bijvoorbeeld de natuurwetenschappen: de impact van publicaties loopt over een veel langere periode door en vertoont veel minder een piek in de eerste paar jaren na verschijnen. Dit betekent dat bibliometrische tellingen met het op de bètawetenschappen geënte gebruikelijke tijdsvenster van slechts enkele jaren een inadequaat beeld geven van de feitelijke impact van publicaties in de geesteswetenschappen.

Naast deze meer theoretische *caveats* is er nog een meer praktische, maar zwaarwegende beperking. Bibliometrische indicatoren geven het – beperkte en, zoals we zagen, zeker interpretatie behoevende – inzicht dat alleen maar als de onderliggende database met publicaties en citaties van een onderzoeker voldoende dekkend is. Dit is lang niet altijd het geval en geldt zeker niet voor de meeste onderdelen van de geesteswetenschappen. De databases van het *Web of Science* van Thomson Reuters, die de basis vormen voor de meeste bibliometrische analyses, bieden geen dekkend overzicht van geesteswetenschappelijke literatuur, voornamelijk omdat boeken en niet-Engelstalige tijdschriften er niet in zijn opgenomen. Onderstaande tabel geeft een overzicht van referenties naar artikelen in *Web of Science*-tijdschriften voor verschillende vakgebieden. In de geesteswetenschappen (onderaan in de tabel) betreft slechts een derde van de referenties artikelen in tijdschriften (de overige referenties gaan naar boeken of bundels) en slechts de helft van die tijdschriften is opgenomen in het *Web of Science*. Meer dan tachtig procent van de referenties in de geesteswetenschappelijke literatuur gaat dus naar publicaties (artikelen of boeken) die niet in het *Web of Science* zijn opgenomen. Zoals uit de tabel blijkt biedt het *Web of Science* vooral voor de biomedische en bètawetenschappen een redelijk goede dekking.

**Tabel 1** Referenties naar artikelen in *Web of Science*-tijdschriften 2002<sup>32</sup>

Vakgebied	Referenties naar publicaties in tijdschriften	Aandeel van WoS-tijdschriften binnen de referenties naar publicaties in tijdschriften	Belang van WoS (kolom1 x kolom2)
Moleculaire biologie en biochemie	96%	97%	92%
Biomedische wetenschappen gericht op de mens	95%	95%	90%
Chemie	90%	93%	84%
Klinische geneeskunde	93%	90%	84%
Theoretische natuurkunde en astronomie	89%	94%	83%
Toegepaste natuurkunde en chemie	83%	89%	73%
Biologische wetenschappen gericht op planten en dieren	81%	84%	69%
Psychologie en psychiatrie	75%	88%	66%
Geologie	77%	81%	62%
Overige sociale wetenschappen / gezondheidswetenschappen	75%	80%	60%
Wiskunde	71%	74%	53%
Economische wetenschappen	59%	80%	47%
Technische wetenschappen	60%	77%	46%
Overige sociale wetenschappen	41%	72%	29%
Geesteswetenschappen	34%	50%	17%

Binnen de geesteswetenschappelijke disciplines varieert het dekkingspercentage tussen 20% (archeologie) en 7% (neerlandistiek en germanistiek).

Een recent onderzoek van het CWTS bij de geesteswetenschappelijke faculteiten van de twee Amsterdamse universiteiten laat een vergelijkbaar beeld zien. Het *Web of Science* mist ook hier het grootste deel van de wetenschappelijke productie van de geesteswetenschappen. In tegenstelling tot de hoge dekkingspercentages voor de bèta- en biomedische wetenschappen van meer dan 85%, komen de geesteswetenschappen niet verder dan respectievelijk 27 en 22%.<sup>33</sup> Citatietellingen gebaseerd op het *Web of Science* geven momenteel dus een onbetrouwbaar beeld voor de geesteswetenschappen.

Op verschillende niveaus wordt inmiddels geprobeerd iets aan die onvolkomenheid van de databases te doen. Thomson Reuters heeft het aantal geesteswetenschappelijke tijdschriften in de afgelopen zes jaren met zo'n 30% vergroot.<sup>34</sup> Sinds 2009 bevat het *Web of Science* ook 1.228 'regionale' tijdschriften, gedefinieerd als 'tijdschriften die zijn gericht op een regionale in plaats van een internationale doelgroep vanwege de focus op regionale thema's vanuit regionaal perspectief'. De helft van die tijdschriften is van Europese herkomst.

32 Moed (2005) 126.

33 *idem*, 130.

34 Moed & al. (2009).



Een alternatief voor *Web of Science* is de database Scopus van Elsevier. De dekingsgraad is hoger dan die van *Web of Science* maar de precieze omvang van de dekking is onduidelijk en adressen van onderzoekers ontbreken, waardoor de database zich moeilijk leent voor consistente analyses. De concurrentie tussen *Web of Science* en Scopus zal op termijn waarschijnlijk leiden tot een betere dekking van geesteswetenschappelijke tijdschriften.<sup>35</sup> Thomson heeft inmiddels op de Frankfurter Buchmesse een citatie-index van boeken aangekondigd, maar het valt te betwijfelen of dit project op afzienbare termijn gerealiseerd zal worden.

Een derde database is die van *Google Scholar*, waarin ook boeken en hoofdstukken in boeken zijn opgenomen. *Google Scholar* is echter niet transparant over de inhoud van de database. Bij bibliometristen en in de onderzoekswereld is niet bekend welke tijdschriften door de zoekmachine worden doorzocht en welke boeken in de databases zijn opgenomen. Daardoor is het ook onmogelijk te weten hoe actueel en omvangrijk zoekresultaten zijn van *Google Scholar*.<sup>36</sup> Als bibliometrische database voor de geesteswetenschappen is ook *Google Scholar* daarom (vooralsnog) niet geschikt.

De conclusies die we kunnen trekken over het gebruik van bibliometrische indicatoren in de geesteswetenschappen zijn al met al niet eenduidig. Enerzijds zijn er geen principiële redenen waarom bibliometrie voor de geesteswetenschappen nooit en nergens zou kunnen werken. In sommige gebieden kunnen bibliometrische indicatoren bruikbare informatie geven. Er zijn echter ook gebieden waar de publicatiecultuur zodanig is dat men met bibliometrie weinig kan beginnen. Daarnaast is er het gesignaleerde probleem van de gebrekkige dekingsgraad van de gebruikte databases, al zal hier de situatie, naar het zich laat aanzien, geleidelijk aan verbeteren. Ten slotte is er het principiële punt dat bibliometrische scores zich niet een-op-een in kwaliteitsoordelen laten vertalen. Er is nog veel onderzoek nodig naar de validiteit en bruikbaarheid van de verschillende bibliometrische instrumenten.

Dit alles betekent dat bibliometrische indicatoren niet overal even bruikbaar zijn en dat ze zeker niet in isolatie als standaardindicator moeten worden gebruikt. Daar waar ze überhaupt toepasbaar zijn, dienen ze altijd te functioneren in samenhang met andere indicatoren en dienen zij gewogen en geïnterpreteerd te worden door *peers*.

## Classificaties van tijdschriften en wetenschappelijke uitgeverijen

Vanwege de problemen bij de toepassing van de meer gangbare vormen van bibliometrie bij onderzoeksbeoordelingen in de geesteswetenschappen wordt op verschillende plekken gewerkt aan bibliografische lijsten, als een alternatief classificatiesysteem voor de output van geesteswetenschappers. Dergelijke bibliografische classificaties kunnen een bibliometrische rol gaan spelen voor zover ze aan evaluatiepanels de mogelijkheid bieden wegingsfactoren te verbinden aan de output van onderzoekers.

---

35 Hicks & Wang (2009).

36 Ibid. Zie ook Harzing (2008).

Enkele jaren geleden nam de European Science Foundation (ESF) het initiatief voor de opstelling van de *European Reference Index for the Humanities (ERIH)*. Het doel van de ERIH was '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'.<sup>37</sup> Deze formulering suggereert dat de ERIH oorspronkelijk ook bedoeld was om als basis te gaan functioneren voor een gedigitaliseerd bibliometrisch instrument, vermoedelijk in de vorm van een database die voor de humaniora een alternatief zou bieden voor het *Web of Science*.

De ESF heeft voor vijftien disciplines werkgroepen van experts gevraagd lijsten te maken met tijdschriften die van belang zijn als publicatiekanalen voor de Europese beoefening van de geesteswetenschappen. Om in de ERIH-lijst te komen moet een tijdschrift als 'good scientific journal' kunnen worden aangemerkt en een objectief vaststelbare *peer review policy* hebben. Daarmee wordt de plaatsing van een tijdschrift op de ERIH-lijst (evenals opname in de *ISI reference index*) een kwaliteitsindicator op zich. Elk tijdschrift krijgt binnen een discipline bovendien een A-, B- of C-status.<sup>38</sup> Momenteel bevatten de vijftien lijsten samen circa 5.200 tijdschriften.<sup>39</sup>

Of de ERIH kan uitgroeien tot een Europese bibliometrische tegenhanger van de *Web of Science*-indices valt te betwijfelen. Het is zelfs de vraag of de lijst ooit zal functioneren als een algemeen geaccepteerd bibliografisch instrument en of de gehanteerde classificaties ingang zullen vinden als kwaliteitsindicatoren. Op de wijze waarop de lijst tot stand is gekomen is veel kritiek geuit. Uit verschillende bronnen, waaronder gesprekken die de commissie voerde met Nederlandse onderzoekers die bij de totstandkoming van de ERIH-lijst betrokken waren, blijkt dat de opname van bepaalde tijdschriften op de lijst te arbitrair was. Aanvankelijk was er sprake van een eenzijdige focus op Angelsaksische tijdschriften, maar later keerde het tij en kwam de discussie juist in het teken te staan van de meer politieke kwestie van de representatie van de grotere *versus* de kleinere taalgebieden. Daarbij werden volgens velen te veel nationale tijdschriften van twijfelachtig of tenminste subtopniveau te hoog ingeschaald op voordracht van commissieleden uit de desbetreffende landen. Ook de controleerbaarheid van de *ranking* door individuele commissieleden liet te wensen over. Er werd immers zwaar geleund op het oordeel van telkens maar één commissielid per onderzoeksdomein. De indeling in A-, B- en C-tijdschriften werd, mede hierdoor, na verloop van tijd te veel een onderwerp van discussie (en is dat nog steeds).

---

37 *European Reference Index for the Humanities* (online).

38 De categorieën worden als volgt gedefinieerd. 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 Daarvan tellen 360 mee in twee disciplines, 33 in drie disciplines en vijf in vier disciplines. De lijst bevat ook Nederlandstalige tijdschriften.

## Nationale classificatiesystemen

De meeste vakgebieden hebben de ERIH-lijst inmiddels terzijde geschoven. Uit een rondvraag blijkt dat de ERIH-lijst in Nederland nauwelijks wordt gebruikt. Uit het bovenstaande blijkt al dat een belangrijke oorzaak van het mislukken van het ERIH-initiatief gezocht moet worden in het feit dat er gestreefd werd naar één Europese lijst, waarin alle taalggebieden en tijdschriften vertegenwoordigd waren. Misschien wilde men dus te veel in één keer. Daarom, en ook vanwege het ontbreken van een overkoepelend bibliometrisch of bibliografisch databestand, suggereren Hicks & Wang<sup>40</sup> (2009) in een opdrachtonderzoek van een consortium van Europese onderzoeksfinanciers de ontwikkeling van nationale systemen.

Onafhankelijk van deze aanbeveling is de ontwikkeling van nationale classificaties van publicatiemedia al in verschillende landen in gang gezet, waaronder Noorwegen, Denemarken en België (Vlaanderen). In Noorwegen heeft de Norwegian Association of Higher Education Institutions in 2004 een nationale onderzoeksdatabase opgezet, die de bibliografische gegevens bevat van alle significante publicaties (*academic publications*) in alle wetenschapsgebieden, dus inclusief de geesteswetenschappen.<sup>41</sup> De database wordt gebruikt om nationale onderzoeksmiddelen te verdelen over wetenschappelijke instellingen. De binnen het systeem erkende publicatietypes zijn artikelen in tijdschriften, artikelen in bundels en monografieën. De publicatiekanalen worden geïnclassificeerd als behorend tot het basisniveau (*level 1*) of tot een select niveau (*level 2*).<sup>42</sup> Dit leidt tot een puntenwaardering op twee niveaus: een monografie krijgt 5 punten op niveau 1 en 8 op niveau 2; voor een artikel in een tijdschrift worden respectievelijk 1 of 3 punten toegekend, en voor een artikel in een bundel 0,7 of 1. Omdat de criteria voor het toekennen van het label *level 2* over de verschillende wetenschapsgebieden heen variëren, kent het systeem drie typen vakgebieden of *academic field groups*, met elk hun eigen set criteria. Het classificatiesysteem is ingevoerd in 2005 en voor het eerst in het allocatiemodel gebruikt in 2006. Het is een vorm van gewogen tellen van publicaties en maakt geen gebruik van citatiescores.<sup>43</sup>

Om vergelijkbare redenen, namelijk het verdelen van onderzoeksgeld over instellingen, is in Vlaanderen in 2009 begonnen met de opbouw van een Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen (VABB SHW).<sup>44</sup> In het nieuwe financieringsstelsel van 2008 gaat het hier om het 'variabele onderzoeksluik' van 28% van de aan de Vlaamse universiteiten toe te delen middelen.

---

40 Hicks & Wang (2009).

41 Norwegian Association of Higher Education Institutions (2004). 'Locale' publicaties, gedefinieerd als publicaties waarvan meer dan tweederde van de bijdragen uit hetzelfde instituut komen, worden niet meegenomen in het systeem.

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

43 Voor een bespreking van het Noorse model, zie Sivertsen (2010) 22-28.

44 Zie voor een uitvoerige beschrijving Engels et al. (2008).

Hiervan wordt 30% bepaald door de onderzoeksoutput. Aanvankelijk wilde men dit deel verdelen volgens de verdeelsleutel van het Bijzonder Onderzoeksfonds (BOF), die sinds 2003 *de facto* gebaseerd is op de output (en citaties) van voornamelijk de biomedische en natuurwetenschappen, omdat men zich wenste te baseren op de databases van het *Web of Science*. Die databases dekken, zoals eerder aangegeven, slechts een relatief klein deel van de publicaties in de geesteswetenschappen en de sociale wetenschappen. Om deze wetenschappen uiteindelijk toch op een eerlijker manier in het model mee te kunnen wegen, is het VABB SHW opgericht. Het VABB moet op basis van kwalitatieve minimumeisen een selectie bevatten van *peer reviewed* academische publicaties van onderzoekers aan Vlaamse universiteiten uit de sociale en humane wetenschappen, die niet zijn opgenomen in het *Web of Science* (tijdschriftartikelen, boeken/boekbijdragen, bijdragen in *proceedings*), maar die wel als output meegewogen moeten worden in het verdeelmodel.<sup>45</sup> Een speciale commissie voor de sociale en geesteswetenschappen ('Gezaghebbend Panel') bepaalt welke publicaties (uit een compleet overzicht van de publicaties van de Vlaamse universiteiten) worden toegelaten tot het VABB en daarmee dus ook welke publicaties meetellen voor het financieringsmodel. Het gaat om een beperkt aantal typen publicaties: tijdschriftartikelen, boeken, *edited volumes*, bijdragen aan boeken en artikelen in *proceedings*.<sup>46</sup> De classificatie is binair: publicaties tellen wel of niet mee; een verdere *ranking* blijft achterwege. Het bestand bevat alleen de referenties (bibliografische beschrijvingen) van de publicaties, niet de publicaties zelf en kan dus ook niet gebruikt worden voor citatiemetingen. Het bestand wordt jaarlijks geactualiseerd en eens per drie jaar vindt een toetsing plaats van de werkwijze en criteria door internationale experts.

Het op deze wijze classificeren van publicaties op basis van een classificatie van tijdschriften en uitgeverijen als kwaliteitsindicator is voor kritiek vatbaar. Ook hier is er bijvoorbeeld niet noodzakelijk een een-op-een-relatie tussen het kwaliteitsniveau van een publicatie en de status (bijvoorbeeld A, B, of C) van het publicatiemedium, in die zin dat ook een publicatie in een B- of C-tijdschrift van hoge kwaliteit kan zijn, temeer daar sommige van deze tijdschriften niche-tijdschriften zijn, waarin vrijwel alle publicaties over een bepaald onderwerp verschijnen. Toch concludeert de commissie dat dit soort classificaties een nuttige rol kan spelen voor zover ze globaal reliëf aanbrengen in de onderzoeksoutput van een onderzoeker of onderzoeksgroep: het feit dat iemand in staat is om met een zekere regelmaat in A-tijdschriften te publiceren mag gezien worden als een teken van kwaliteit. Het toekennen van een wegingsfactor aan publicaties in A-, B- en C-tijdschriften en het gebruik daarvan in een evaluatief puntenstelsel suggereert naar de mening van de commissie echter een exactheid die deze

---

45 De criteria voor toelating zijn vergelijkbaar met die in Noorwegen: publiek toegankelijk zijn, identificeerbaar via ISBN of ISSN, een bijdrage leveren aan de ontwikkeling van nieuwe inzichten, beoordeeld zijn in een *peer review*-proces, toeschrijfbaar zijn aan een discipline van de sociale en humane wetenschappen.

46 De mogelijkheid wordt opengelaten in een later stadium ook andere typen producten zoals tekstedities, catalogi en databestanden op te nemen.

specifieke kwaliteitsindicator niet kan bieden. De commissie is dan ook van mening dat de classificatie in A-, B- en C- tijdschriften binnen een systeem van kwaliteitsindicatoren in onderzoeksrapportages en onderzoeksbeoordelingen wel een relevante rol dient te spelen, maar dat de weging geen automatische calculatie moet zijn, en overgelaten moet worden aan het deskundige oordeel van *peers*.

De vraag of men dit soort classificaties – al dan niet met puntentellingen – wil gebruiken als basis voor de allocatie van middelen, zoals in Noorwegen en Vlaanderen gebeurt, is nog weer een verdergaande, politieke vraag, die de commissie niet heeft willen beantwoorden. Wel is zij van mening dat juist waar de gedigitaliseerde bibliometrie (citatiescores, H-index, et cetera) in veel gebieden van de geesteswetenschappen onvoldoende bruikbaar is, een classificatiesysteem van publicatiemedia (zowel tijdschriften als uitgeverijen en boekseries) nuttige alternatieve indicatoren zou kunnen leveren, mits met beleid toegepast en in het besef van de hierboven genoemde beperkingen.

De commissie is van mening dat dergelijke lijsten van A-, B- en C-tijdschriften vooralsnog het beste op nationaal niveau kunnen worden samengesteld, waarbij overigens de initiatieven in andere landen als nuttige *benchmark* kunnen fungeren. Daarnaast meent de commissie dat in Nederland een dergelijke taak zou moeten worden uitgevoerd door disciplinespecifieke gremia, waarbij in eerste instantie gedacht kan worden aan de landelijke onderzoekscholen. De lijsten moeten bovendien van tijd tot tijd door deze zelfde disciplinaire gremia worden herijkt. Dit proces van classificatie zal regie vragen. De commissie is van mening dat de KNAW, gezien haar onafhankelijke positie, haar gezag en haar betrokkenheid bij landelijke onderzoekscholen deze regierol goed kan vervullen. Een inventarisatie van de mogelijkheden om voor enkele geselecteerde vakgebieden tot een dergelijke classificatie te komen zal deel uitmaken van de pilotfase van het onderhavige project (zie hoofdstuk 5).

## Conclusies

Bibliometrische kwaliteitsindicatoren dienen met voorzichtigheid te worden gehanteerd. Tellingen van artikelen meten slechts productiviteit; tellingen van citaties meten impact, die niet noodzakelijk identiek is aan kwaliteit en vaak ook afhankelijk is van de publicatie- en citatiecultuur in een bepaald vakgebied of deel daarvan. Zij doen dat bovendien uitsluitend in die gebieden die in hoge mate worden gedekt door het *Web of Science* en andere relevante databases. Voor de geesteswetenschappen is de situatie nog zeer onbevredigend: een te grote hoeveelheid boekpublicaties en niet-Engelstalige artikelen wordt nog steeds buiten beschouwing gelaten, waardoor citatietellingen weinig kunnen zeggen. Zodra voor bepaalde geesteswetenschappelijke disciplines de gebruikte databases een adequaat dekkingsniveau hebben bereikt, zal het gebruik van bibliometrische indicatoren meer in beeld komen. Ook dan moet echter gewaakt worden voor simplistische visies op hun bruikbaarheid. Bibliometrische indicatoren dienen altijd gezien te worden in samenhang met andere indicatoren en binnen de kwalitatieve context van *peer review*.

Als alternatief ondersteunt de commissie de ontwikkeling van lijsten van A-, B- en C-tijdschriften voor verschillende disciplinegebieden als middel om een minimum aan vergelijkbaarheid tussen verschillende vakgebieden te bereiken. Maar ook hier past de kanttekening dat deze lijsten met beleid en binnen het kader van een evenwichtige *peer review*-procedure dienen te worden toegepast. Bovendien zullen tijdschriftenlijsten dynamisch moeten zijn. Zij worden voor een beperkte tijdsspanne vastgesteld en zullen periodiek door een gezaghebbend orgaan moeten worden herijkt.

#### RANDVOORWAARDEN

Voor een systeem van kwaliteitsindicatoren leveren de beschouwingen van dit hoofdstuk twee nieuwe randvoorwaarden op:

- bibliometrische kwaliteitsindicatoren moeten met voorzichtigheid worden gebruikt. Waar überhaupt toepasbaar moeten zij functioneren binnen een systeem van onderzoeksbeoordeling via *peer review*: weging en interpretatie moet door *peers* en andere experts gebeuren en kan geen kwestie zijn van automatische calculatie;
- een systeem van kwaliteitsindicatoren voor de geesteswetenschappen moet een beroep kunnen doen op een breed geaccepteerde *ranking* van publicatiemedia (tijdschriften en boeken of reeksen).

# 5. NAAR EEN SYSTEEM VAN KWALITEITSINDICATOREN VOOR DE GEESTESWETENSCHAPPEN

In dit slothoofdstuk worden om te beginnen de randvoorwaarden voor een passend systeem van kwaliteitsindicatoren voor de geesteswetenschappen nog eens uiteengezet. Deze randvoorwaarden zijn gebaseerd op de conclusies van de vorige hoofdstukken. Een groot deel van deze randvoorwaarden wordt vervolgens verwerkt in een schets van een systeem van kwaliteitsindicatoren. De overige randvoorwaarden worden ten slotte verwerkt in aanbevelingen voor de toepassing en implementatie van dit systeem en in richtlijnen voor de pilotfase van dit project.

## Randvoorwaarden

In de voorgaande hoofdstukken werd de globale context duidelijk waarbinnen een systeem van kwaliteitsindicatoren zou moeten functioneren. Kwaliteitsindicatoren zijn, zo werd aangegeven, nodig om het proces van *peer review* te versterken en van een externe basis te voorzien (hoofdstuk 1). Een *systeem* van kwaliteitsindicatoren zou het proces van *peer review* ook moeten vergemakkelijken door het te structureren (hoofdstuk 3). Zo'n systeem moet zowel output-indicatoren als *esteem*-indicatoren bevatten en indicatoren voor zowel wetenschappelijke als maatschappelijke kwaliteit. Het dient zo breed te zijn als de wetenschapspraktijk vereist, maar voldoende flexibel om in verschillende contexten maatwerk te kunnen bieden, en het dient ruimte te bieden voor aanpassingen of toevoegingen indien de verdere digitalisering van de wetenschappelijke productieprocessen zulks nodig maakt (hoofdstuk 1).

We hebben verder betoogd dat de geesteswetenschappen in vergelijking met andere wetenschapsgebieden niet vragen om een andere benadering van kwaliteit, en dus ook niet om een ander *type* kwaliteitsindicatoren. Wel vragen de geesteswetenschappen – maar daarin verschillen ze niet van de meeste andere wetenschapsgebieden – om een vrij breed scala van indicatoren voor kwaliteit dat recht doet aan de diversiteit aan producten, doelgroepen, en publicatieculturen die men binnen het gebied aantreft. Dit betekent dat het primaat niet kan liggen bij bibliometrische indicatoren die gebaseerd zijn op databestanden die voornamelijk uit Engelstalige tijdschriftpublicaties bestaan. Daarnaast moet het systeem ruimte bieden voor andere vormen van output dan wetenschappelijke publicaties, zoals databases, catalogi en tekstedities (hoofdstuk 2).

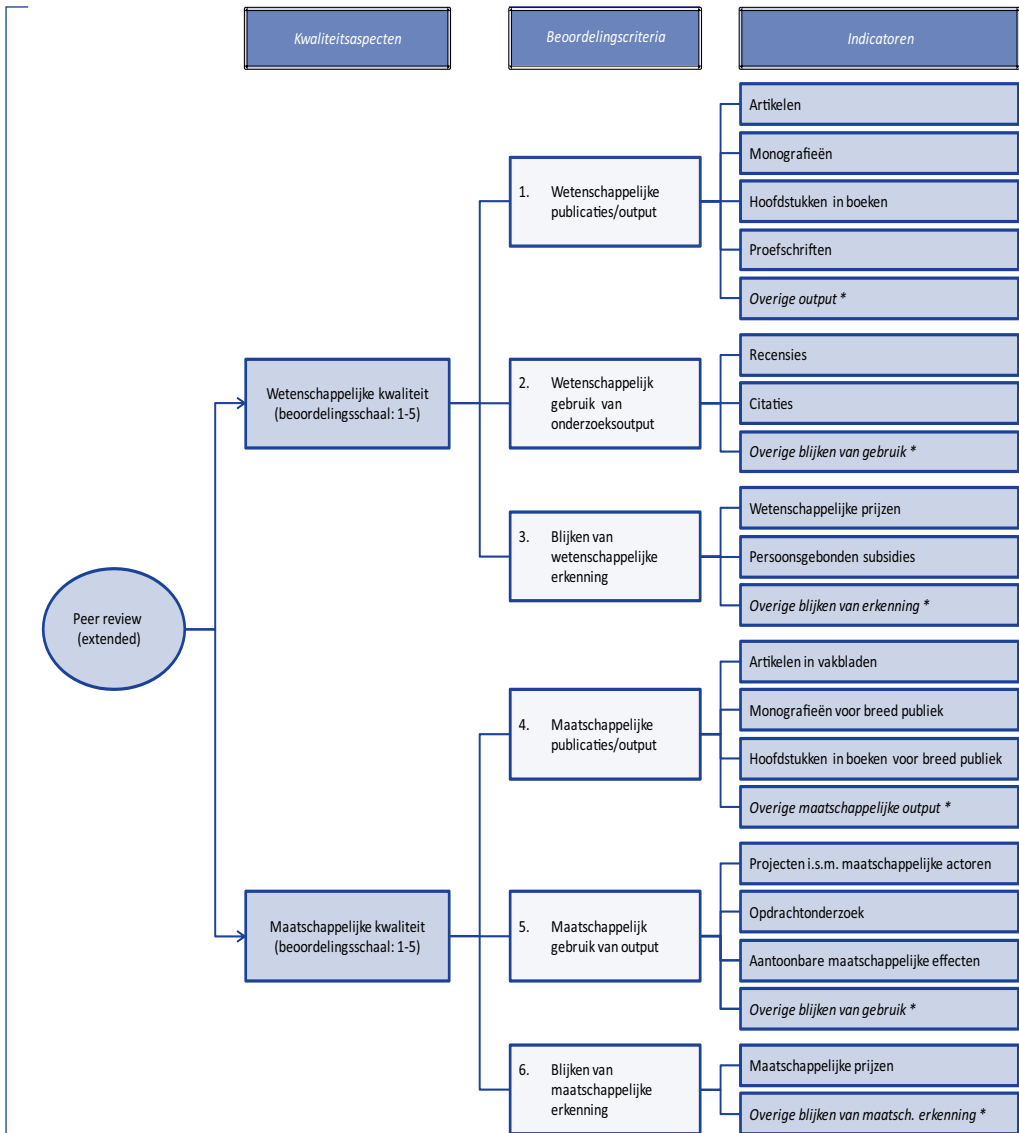
*Peer review* en indicatoren, zo is verder betoogd, staan in een zekere wisselwerking tot elkaar: indicatoren ondersteunen *peer review*, maar voor de specifieke weging van de verschillende indicatoren is het *expert judgement* van de *peers* noodzakelijk. Dit geldt in de huidige omstandigheden in het bijzonder, maar zeker niet uitsluitend, voor de bibliometrische indicatoren. Geïsoleerd en ongerefleeteerd gebruik van bibliometrische indicatoren is onverantwoord in het algemeen en in de geesteswetenschappen in het bijzonder. Om reliëf aan te brengen in de output-indicatoren is het ten slotte volgens de commissie gewenst dat er een landelijk gedragen classificatie komt van geesteswetenschappelijke tijdschriften, uitgeverijen en boekseries. Ook deze classificatie dient echter vooral als input voor het proces van kwalitatieve *peer review*, niet als basis voor automatische calculaties (hoofdstukken 3 en 4).

Bij sommige *stakeholders* heeft de gedachte postgevat dat een simpel systeem – in de zin van een systeem met enkele basisindicatoren dat de kwalitatieve weging door *peers* in sterke mate objectiviseert of zelfs vervangt – wenselijk en haalbaar is. Volgens de commissie is dat niet het geval. Een adequaat systeem is niet simpel en een beperkt systeem is niet adequaat. Wel heeft de commissie gezocht naar een systeem dat maximaal eenvoudig is (in de zin van: werkbaar) door zijn flexibiliteit, en heeft zij geprobeerd in haar aanbevelingen voor implementatie de praktische toepasbaarheid en de doelstelling van het vermijden van extra bureaucratische ballast steeds voor ogen te houden (hoofdstuk 3).

## Schets van een systeem van kwaliteitsindicatoren voor de geesteswetenschappen

Een systeem van kwaliteitsindicatoren voor de geesteswetenschappen dat rekening houdt met de bovenstaande overwegingen en randvoorwaarden stelt de commissie zich als volgt voor:





*Figuur 1* schema van criteria en indicatoren  
 (\* disciplinespecifieke/contextspecifieke indicatoren)

Dit systeem en de manier waarop het zou moeten worden toegepast verdienen op een aantal punten nadere toelichting. Wij volgen het schema van links naar rechts.

- Bij de beoordeling van geesteswetenschappelijk onderzoek kan onderscheid worden gemaakt naar wetenschappelijke en maatschappelijke kwaliteit. De commissie stelt zich voor dat peers, indien nodig ondersteund door externe experts op het

gebied van maatschappelijke kwaliteit of door maatschappelijke stakeholders,<sup>47</sup> beide aspecten standaard beoordelen op een vijfpuntsschaal, die overeenkomt met het Standard Evaluation Protocol.

- *Peers*, indien nodig ondersteund door externe experts op het gebied van maatschappelijke kwaliteit of door maatschappelijke *stakeholders*, beoordelen deze aspecten aan de hand van de in de tweede kolom gespecificeerde beoordelingscriteria. Voor zowel wetenschappelijke als maatschappelijke kwaliteit zijn dat: output, aantoonbaar gebruik van output en blijken van erkenning (*esteem*). Deze beoordelingscriteria zijn datgene waar de *peers* 'iets van moeten vinden'. Geaggregeerd leveren deze oordelen de eindoordelen op (volgens vijfpuntsschaal) op de twee hoofdaspecten (linkerkolom).
- In de rechterkolom vindt men een lijst van indicatoren per criterium. De te visiteren disciplines, instellingen, instituten of programma's bepalen zelf welke indicatoren ze wel en niet gebruiken (welke relevant zijn voor de specifieke discipline, dan wel welke relevant zijn in verband met de specifieke missie van een instituut of programma of de taakomschrijving van individuele onderzoekers).<sup>48</sup> De commissie heeft geen basisset van indicatoren gespecificeerd, omdat de prescriptieve afgrenzing van wat tot die basisset zou moeten behoren hoe dan ook arbitrair zou zijn, terwijl te verwachten is dat in de praktijk vanzelf duidelijk wordt welke indicatoren (zoals wetenschappelijke artikelen en monografieën) in verreweg de meeste contexten relevant geacht zullen worden.
- Er moeten richtlijnen worden geformuleerd voor de levering van de relevante gegevens. Men kan zich bijvoorbeeld voorstellen dat men in het geval van wetenschappelijke artikelen het volgende vraagt: complete bibliografische gegevens, aantallen publicaties in A-, B- en C-media en een nader te bepalen aantal *sample*-publicaties. De commissie wil de pilotfase van dit project gebruiken om werkbare richtlijnen te ontwerpen. Voor een eerste schematische uitwerking (die als input kan fungeren voor de pilotfase), zie bijlage 1.

Dit systeem van kwaliteitsindicatoren voldoet aan de in eerdere hoofdstukken genoemde randvoorwaarden. Het omvat zowel wetenschappelijke als maatschappelijke kwaliteit en zowel output- als *esteem*-indicatoren. Het is breed en flexibel: het geeft de mogelijkheid om, waar dat geboden is, bepaalde indicatoren, zoals bibliometrische indicatoren, buiten beschouwing te laten en bij elk van de zes beoordelingscriteria kan een categorie 'overige indicatoren' vakspecifiek of contextspecifiek worden ingevuld. Ook de toevoeging van nieuwe indicatoren op basis van een voortschrijdende digitalisering van de wetenschappelijke productieprocessen is mogelijk. Het systeem is dus

---

47 Voor dit soort vormen van *extended peer review*, zie hoofdstuk 3; zie ook de opmerkingen hieronder over de pilotfase van dit project.

48 Bij het benadrukken van het belang om ruimte te scheppen voor missiegerelateerde kwaliteitsindicatoren sluit de commissie aan bij de conclusies van het KNAW-rapport *Judging Research on its Merits* uit 2005.

niet in beton gegoten. Het biedt verder de mogelijkheid van een kwantitatieve invulling van sommige indicatoren (aantallen publicaties, bibliometrische kengetallen) maar de uiteindelijke weging van deze criteria ten behoeve van een eindoordeel over de kwaliteitsaspecten (linkerkolom) is een kwalitatieve en komt toe aan *peers*.

## Contexten van toepassing

Zoals in hoofdstuk 1 aangegeven, vinden kwaliteitsbeoordelingen plaats in verschillende contexten en met verschillende oogmerken. In het onderstaande wordt kort ingegaan op de manier waarop het bovenstaande schema zou kunnen functioneren in drie standaardcontexten: (a) periodieke visitaties, (b) beoordelingen van onderzoeksvorstellen en individuele cv's, en (c) prestatiebeoordelingen van individuele personeelsleden in faculteiten en instituten.

### (a) Onderzoeksvisitaties in het kader van het SEP-protocol

Eens in de zes jaar worden publiek gefinancierde onderzoeksinstituten en hun onderzoeksprogramma's extern geëvalueerd. Voor deze evaluatie is door de VSNU, NWO en KNAW het *Standard Evaluation Protocol* (SEP) opgesteld. Sommige kwaliteitscriteria in dit protocol (originaliteit, samenhang van een programma; zie SEP-protocol, criterium A1) kunnen alleen direct door *peers* worden beoordeeld. Ten aanzien van een aantal andere criteria voor wetenschappelijke kwaliteit (bijdrage aan het veld, output, relevantie), maar ook ten aanzien van criteria als academische reputatie (zie SEP-protocol, criterium A4), productiviteit (zie SEP-protocol onder B2) en maatschappelijke relevantie (zie SEP-protocol, criterium C) kan *peer review* worden ondersteund door een relevante selectie uit het systeem van kwaliteitsindicatoren. Indicatoren voor de maatschappelijke relevantie van onderzoek kunnen desgewenst verder worden verfijnd aan de hand van de handreiking die recentelijk is verschenen in het kader van het ERiC-project.<sup>49</sup>

Op dit moment is er in de geesteswetenschappen nog geen uniforme praktijk van landelijke, disciplinegebonden visitaties. Waar disciplinegebonden visitaties plaatsvinden kan per discipline door een gezaghebbend orgaan (een disciplineoverlegorgaan of een landelijke onderzoeksschool) een relevant geachte minimumselectie worden gemaakt. Waar dit niet het geval is, doet het instituut dat de opdracht tot visitatie geeft dit. Daarnaast kunnen individuele instituten of programma's missiegerelateerde indicatoren specificeren.

Om inhoud te geven aan de in het SEP genoemde categorisering van tijdschriften in een top 10% en een top 25% (A- en B-categorie) is de eerder genoemde landelijk gedragen classificatie van publicatiemedia naar de mening van de commissie onontbeerlijk.

---

49 ERiC (2010)

## **(b) Beoordelingen van onderzoeksvorstellen en individuele cv's**

Bij onderzoeksvorstellen bij NWO en vergelijkbare organisaties worden referenten en juryleden geacht een oordeel te kunnen geven over zowel de kwaliteit van het voorstel als de wetenschappelijke kwaliteit van de indiener.<sup>50</sup> Het eerste aspect is prospectief en kan dus alleen direct door *peers* (referenten) worden beoordeeld. Het laatste aspect is retrospectief en komt in feite neer op het beoordelen van een onderzoeks-cv van de aanvrager. Het geschetste systeem van kwaliteitsindicatoren kan hier als referentiekader gelden. Dat geldt ook voor contexten waarin sprake is van toekenning van subsidies die gelden als prijs of erkenning voor wetenschappelijke excellentie (bijvoorbeeld Spinoza-prijzen en Heinekenprijzen). Naast de basisindicatoren zullen in de contextbeoordelingen van dit type vooral *esteem*-indicatoren een rol spelen.

Een probleem dat door NWO en andere subsidiënten wordt gesignaleerd is de gebrekkige vergelijkbaarheid van verschillende vakgebieden, waardoor de allocatie van middelen over die vakgebieden iets willekeurig lijkt te krijgen. In beginsel zou dit probleem op te lossen zijn door één of enkele kwaliteitsindicatoren als gemeenschappelijke meetlat te gebruiken. De commissie is van mening dat een dergelijke oplossing zich niet aandient. Hoewel, zoals in hoofdstuk 2 betoogd is, verschillende wetenschapsgebieden min of meer dezelfde notie van kwaliteit hanteren, vraagt de diversiteit aan publicatieculturen om een variëteit aan disciplinespecifieke kwaliteitsindicatoren. Een simpel en beperkt systeem dat gemakkelijke vergelijkingen mogelijk maakt is een illusie.

Dit wil niet zeggen dat er geen enkele vorm van compatibiliteit tussen disciplines is waar het gaat om kwaliteitsbeoordelingen. Een zekere mate van vergelijkbaarheid binnen de geesteswetenschappen kan bijvoorbeeld bereikt worden door *across the board* te gaan werken met classificaties van tijdschriften en boekseries, zoals in dit rapport wordt aanbevolen. Daarnaast zou men op termijn kunnen werken met bibliometrische indicatoren die genormeerd zijn per vakgebied. Dit lijkt echter momenteel prematuur, en is hoe dan ook omslachtig en nog altijd tamelijk grofmazig, omdat de publicatie- en citatieculturen ook binnen vakgebieden aanzienlijk kunnen verschillen.

## **(c) Prestatiebeoordelingen binnen faculteiten en instituten**

In praktisch alle geesteswetenschappelijke faculteiten en instituten worden onderzoeksgroepen en onderzoekers jaarlijks beoordeeld op wetenschappelijke kwaliteit en productiviteit. Deze beoordelingen vinden vaak, maar niet uitsluitend, plaats in R&O-gesprekken in het kader van het personeelsbeleid. Centraal staat hierbij doorgaans de beoordeling van de onderzoeksoutput in relatie tot de beschikbare onderzoekstijd, die consequenties kan hebben voor de allocatie of herallocatie van onderzoekscapaciteit en dus voor het takenpakket van de individuele medewerker. Ook hier kan het systeem

---

<sup>50</sup> Het beoordelingskader voor aanvragen in de NWO Vernieuwingsimpuls kent bijvoorbeeld ook drie beoordelingsaspecten: *assessment of quality of the researcher, innovative character of the proposal, assessment of quality of the research proposal*.

van kwaliteitsindicatoren worden toegepast. De evaluatie is echter allicht minder breed dan bij de bovengenoemde contexten (a) en (b). Doorgaans zal men zich kunnen beperken tot enkele output-indicatoren. Steeds meer geesteswetenschappelijke faculteiten en onderzoeksinstituten gebruiken een puntensysteem, of overwegen de invoering daarvan, om het proces van personeelsbeoordeling en de verdeling van onderzoekstijd te structureren. In de meeste geesteswetenschappelijke faculteiten wordt van onderzoekers verlangd dat zij per 0,2 fte onderzoeksinzet gemiddeld over een aantal jaren minimaal één wetenschappelijk tijdschriftartikel per jaar leveren. Boeken krijgen daarbij een waarde toegekend die, afhankelijk van de context, gelijk staat aan die van een aantal tijdschriftartikelen. De eerder genoemde landelijke classificatie van tijdschriften, uitgeverijen van boeken en boekseries per discipline zal helpen een globaal beeld te verkrijgen van het speelveld waarop een individuele onderzoeker zich begeeft.

## Vervolgacties en aanbevelingen

Met de analyses van dit rapport en de daaraan gekoppelde schets van een systeem van kwaliteitsindicatoren voor de geesteswetenschappen is de in de inleiding geschetste 'conceptuele fase' van het werk van de commissie afgerond. Deze fase wordt, zoals aangekondigd, gevolgd door een experimentele pilotfase waarin het systeem op zijn bruikbaarheid wordt onderzocht en eventueel aangepast of verfijnd, terwijl er tevens aanbevelingen zullen worden geformuleerd voor implementatie.

Deze pilotfase van het traject richting eindrapport zal de volgende elementen omvatten.

1. Het CWTS heeft in opdracht van het CvB van de Erasmus Universiteit (EUR) een onderzoek gestart naar de publicatiecultuur binnen het vakgebied Geschiedenis. Daartoe heeft men de outputgegevens in METIS van de Nederlandse universiteiten opgevraagd. Bibliometristen onderzoeken hoe groot daarbinnen de ISI-dekking is en gaan vervolgens in overleg met historici van de EUR om te kijken of en hoe het resterende 'gat' kan worden gedicht en of er een min of meer complete database gevormd kan worden die als basis kan dienen voor bibliometrisch onderzoek. De commissie wil dit project in nauw overleg met het CWTS monitoren omdat het de mogelijkheden in kaart brengt om met verbeterde middelen bibliometrisch onderzoek te doen binnen een discipline. De invalshoek van de commissie is daarbij een neutrale. Zij committeert zich noch bij voorbaat aan nut of noodzaak van een dergelijke operatie, noch aan de beleidsmatige doelstellingen die de EUR hiermee heeft of zou hebben. Centraal staat voor de commissie de feitelijke vraag wat deze exercitie oplevert, of deze navolging verdient en of de eventueel gecreëerde meerwaarde in verhouding staat tot de investeringen in geld en menskracht.
2. De Commissie Kwaliteitsindicatoren Geesteswetenschappen zal de regie voeren over twee concrete pilotstudies. Het gaat om proefevaluaties bij het Meertens

Instituut van de KNAW en bij het Instituut voor Cultuurwetenschappelijk Onderzoek Groningen (ICOG) van de Rijksuniversiteit Groningen. In beide pilots zal het hier ontwikkelde systeem van kwaliteitsindicatoren worden getest, en zal in het bijzonder bekeken worden

- of het systeem in de praktijk voldoende differentiërend is;
- wat de mogelijkheden en problemen zijn bij het opstellen van disciplinespecifieke classificaties van publicatiekanalen;
- hoe een maximaal informatief maar minimaal belastend systeem kan worden ontwikkeld voor de levering van het materiaal bij onderzoeksbeoordelingen (bijvoorbeeld een modulair systeem dat in verschillende evaluatiecontexten bruikbaar is) en hoe dit zich verhoudt tot de categorieën binnen het bestaande registratiesysteem METIS;
- wat de mogelijkheden en onmogelijkheden zijn van een verdere verfijning van de indicatoren en de beoordelingsprocedure op het aspect van maatschappelijke kwaliteit (bijvoorbeeld door het bij het proces betrekken van maatschappelijke *stakeholders*), waarbij de aanbevelingen van de *Handreiking Evaluatie van Maatschappelijke Relevantie* van ERiC (Evaluating Research in Context) uit 2010 leidend zijn.

De bevindingen van de pilotfase zullen worden verwerkt in het eindrapport van de commissie. Vooruitlopend op deze eindrapportage meent de commissie op basis van het huidige conceptrapport reeds de contouren te kunnen aangeven van twee aanbevelingen aan de KNAW:

- ten eerste zou de KNAW het voortouw dienen te nemen in, en vervolgens de regie moeten voeren over, een landelijk initiatief om te komen tot een breed gedragen disciplinegeoriënteerde classificatie van wetenschappelijke tijdschriften en boekseries voor de geesteswetenschappen. Het ligt voor de hand hierbij in eerste instantie de landelijke onderzoekscholen in te schakelen, die immers kunnen gelden als platforms van deskundigheid voor hun respectievelijke disciplines;
- ten tweede heeft de commissie kunnen constateren dat haar bevindingen op een aantal cruciale punten overeenkomen met die van het onlangs verschenen *Rapport Kwaliteitsbeoordeling in de Ontwerpde en Construerende Disciplines*. Dit maakt de vraag opportuun hoe de situatie in andere wetenschapsgebieden is. De KNAW zou een discussie over kwaliteitsindicatoren kunnen agenderen binnen haar Raden voor de diverse wetenschapsgebieden om eventuele verschillende opvattingen over kwaliteitsindicatoren in kaart te brengen en de mogelijkheden van verdere consensusvorming te onderzoeken.

# BRONNEN

- 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*. Den Haag: AWT.
- AWT (2007), *Alfa en gamma stralen. Valorisatiebeleid voor de alfa- en gammawetenschappen*, Den Haag: 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](http://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](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 et 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), *Kwaliteitszorg in de wetenschap. Van SEP naar KEP: balans tussen rechtvaardigheid en eenvoud.*
- 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* (Engelse vertaling).
- 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 en NWO (2009), *Standard Evaluation Protocol 2009-2015*, [www.knaw.nl/sep](http://www.knaw.nl/sep)
- Waltman, L. & N. J. van Eck (2009), 'A simple alternative to the H-index', *ISSI-Newsletter*, vol. 05 / nr. 3, September 2009.



# BIJLAGE 1

## OVERZICHT INDICATOREN (INPUT VOOR PILOTFASE)

Beoordelings-criteria	Indicatoren	Informatiebehoefte (nader te preciseren in pilotfase)
1. Wetenschappelijke publicaties	Artikelen	Lijst Aantallen, gemarkeerd naar A-, Ben C-tijdschriften Selectie van <i>key publications</i>
	Monografieën	Lijst Aantallen, gemarkeerd naar A-, Ben C-uitgeverijen en series Selectie van <i>key publications</i>
	Hoofdstukken in boeken	Lijst Aantallen, gemarkeerd naar A-, Ben C-uitgeverijen en series Selectie van <i>key publications</i>
	Afgeronde proefschriften	Lijst per onderzoeksgroep Lijst per onderzoeker (1 <sup>e</sup> promotor)
	<i>Overige wetenschappelijke output, bijvoorbeeld congresbundels, redacties, handboeken, woordenboeken, tekstedities, databases, software</i>	Per context nader te bepalen welke kwantitatieve en/of kwalitatieve informatie geleverd moet worden
2. Wetenschappelijk gebruik van onderzoeksoutput	Recensies	Naspeurbare verwijzingen naar <i>samples</i>
	Bibliometrische indicatoren	Citatieanalyse (alleen als de databases voldoende dekkend zijn)
	<i>Overige blijken van gebruik, bijvoorbeeld verkochte/uitgeleende boeken, gedownloade artikelen, gebruik van databases, software</i>	Per context nader te bepalen welke kwantitatieve en/of kwalitatieve informatie aangeleverd moet worden
3. Blijken van wetenschappelijke erkenning	Wetenschappelijke prijzen	Simpele vermelding met jaartallen
	Persoonsgebonden subsidies (VENI, VIDI, VICI, Spinoza, ERC)	Simpele vermelding met jaartallen
	<i>Overige blijken van wetenschappelijke erkenning, bijvoorbeeld lidmaatschappen van academies, eredocoraten, externe benoemingen, uitnodigingen voor keynote-lezingen, uitnodigingen voor handboeken, woordenboeken of tekstedities, chief/full editorships van gerenomeerde (inter)nationale tijdschriften/boeken/buldels, adviseurs-/referentschappen (bijvoorbeeld NWO, visitatiecommissies), fellowships bij prestigieuze (buitenlandse) instituten</i>	Per context nader te bepalen welke kwantitatieve en/of kwalitatieve informatie aangeleverd moet worden

<b>4. Maatschappelijke publicaties</b>	Artikelen in vakbladen (niet primair-wetenschappelijke tijdschriften)	Lijst Selectie van <i>key publications</i>
	Monografieën voor niet-wetenschappelijke professionals en geïnteresseerden	Lijst Selectie van <i>key publications</i>
	Hoofdstukken in boeken voor niet-wetenschappelijke professionals en geïnteresseerden	Lijst Selectie van <i>key publications</i>
	<i>Overige maatschappelijke output, bijvoorbeeld bundels voor niet-wetenschappelijke professionals en geïnteresseerden, redacties van vakbladen, handboeken, woordenboeken, tekstedities, databases, software, tentoonstellingen, catalogi, vertalingen, beleidsadviesrapporten</i>	Per context nader te bepalen welke kwantitatieve en/of kwalitatieve informatie verlangd wordt
<b>5. Maatschappelijk gebruik van onderzoeksoutput</b>	Projecten uitgevoerd in samenwerking met maatschappelijke actoren	Simpele vermelding met jaartallen
	Oprachtonderzoek	Simpele vermelding met jaartallen
	Aantoonbare maatschappelijke effecten van onderzoek	Simpele vermelding met jaartallen
	<i>Overige vormen van maatschappelijk gebruik, bijvoorbeeld recensies, citaties in (beleids)rapporten, gebruik van publicaties, media-aandacht, verkochte/ uitgeleende boeken</i>	Per context nader te bepalen welke kwantitatieve en/of kwalitatieve informatie verlangd wordt
<b>6. Blijken van maatschappelijke erkenning</b>	Maatschappelijke prijzen	Simpele vermelding met jaartallen
	<i>Overige blijken van maatschappelijke erkenning, bijvoorbeeld maatschappelijke benoemingen, uitnodigingen voor lezingen, uitnodigingen voor mediaoptredens, adviesfuncties/lidmaatschappen van adviescommissies</i>	Per context nader te bepalen welke kwantitatieve en/of kwalitatieve informatie verlangd wordt.

# BIJLAGE 2

## ENKELE BUITENLANDSE VOORBEEDEN VAN ONTWIKKELINGEN ROND INDICATOREN VOOR GEESTES- EN SOCIALE WETENSCHAPPEN

In een groot aantal landen wordt nagedacht over het gebruik van indicatoren in de geestes- en sociale wetenschappen. Het rapport van de SWR en de RGW *Judging Research on its Merits* uit 2005 is een voorbeeld uit de Nederlandse context. In dat rapport werd geen systeem voorgesteld, maar wel een richting aangegeven waarin een nieuw systeem van indicatoren zich zou kunnen ontwikkelen. In ons huidige rapport zijn we uitgebreid ingegaan op twee voorbeelden van systemen die al in een vergevorderd stadium van ontwikkeling zijn (uit Noorwegen en uit Vlaanderen). Die zullen we hier niet opnieuw bespreken. Hieronder geven we een aantal voorbeelden van studies en voorstellen die mede ten grondslag hebben gelegen aan de ontwikkeling van het denken binnen de commissie. Het zijn geen uitgewerkte systemen, maar ze geven wel aan waar de belangrijke discussiepunten liggen voor eventueel nieuw te ontwikkelen indicatoren. Het gaat daarbij zowel om indicatoren voor de beoordeling van de wetenschappelijke kwaliteit als om indicatoren voor maatschappelijke impact. De studies worden in het onderstaande telkens eerst kort gepresenteerd waarna een integrale en (ook qua Engels) ongewijzigde samenvatting volgt, ontleend aan het desbetreffende rapport.

1. Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten, *Bibliometrie in de Humane Wetenschappen (2004)*, rapport van de werkgroep bibliometrie in de humane wetenschappen.

Het betreft een rapport van twee commissies die de opdracht hadden de mogelijkheid te onderzoeken van een bibliometrisch systeem voor de geestes- en sociale wetenschappen. Ze kunnen gezien worden als een voorloper van het Vlaams Academisch Bibliografisch Bestand dat we in het rapport bespreken.

### **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

Het betreft een brede studie uitgezet door een aantal *research councils*. Het rapport werd midden 2010 aangeboden aan de opdrachtgevers, maar er is nog geen follow-up bekend. Deze studie liep parallel aan de poging een *European Reference Index for the Humanities* (ERIH) te maken, maar staat daar los van.

***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 out 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 \(2009\).](#)

Het betreft een gezamenlijk project van SCIMago Research Group, CSIC Madrid, University of Granada, Spain, en Centre for Science and Technology Studies (CWTS), Leiden University. Het project onderzoekt de inhoud van een groot aantal bibliografische en bibliometrische databases op hun geschiktheid voor gebruik in de sociale en geesteswetenschappen (onder meer *Web of Science*, *Google* en *Google Scholar*, *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)*, Cybergego, *European journal of Geography*, n°484, 2010

Het betreft een vergelijkend onderzoek naar de dekking van de bekendste databases voor de geesteswetenschappen.

### **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 *Cybergego*. 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 pour l'Evaluation de la Recherche 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 multidisciplinary nature 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 Humanities and Social Sciences*, *Journal of the American Society for Information Science and Technology*, 60(6):1083–1096, 2009

Het betreft een voorstel om een maat te ontwikkelen die iets zegt over boeken. De maat is gebaseerd op de keuzes die bibliothecarissen maken, ervan uitgaande dat zij hun keuze baseren op kennis van de reputaties van onderzoekers en van de belangstelling van potentiële lezers.

### **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. [De Britse Arts and Humanities Research Council \(AHRC\) is al jaren actief om de impact van onderzoek aan te tonen. Hieronder twee voorbeelden van de impact van \(digitale\) kunst op andere vakgebieden.](#)

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

Het betreft een Canadese studie naar de impact van sociale en geesteswetenschappen op de economie. Interessant is om te zien hoe aan de hand van concrete gegevens wordt aangegeven hoe belangrijk onderzoek in de geesteswetenschappen is voor de economie.

### **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 use 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.

# BIJLAGE 3

## LIJST VAN GEÏNTERVIEWDEN

Prof. dr. Hans Bertens	Universiteit Utrecht
Prof. dr. Wim Blockmans	NIAS/Universiteit Leiden
Drs. Annemarie Bos	NWO
Prof. dr. Kees de Bot	Rijksuniversiteit Groningen
Prof. dr. Rosemarie Buikema	Ned. Onderzoeksschool Vrouwenstudies
Prof. dr. Karin Bijsterveld	Onderzoeksschool WTMC
Drs. Liesbeth Cilek	Universiteit Utrecht
Prof. dr. Koenraad Debackere	Katholieke Universiteit Leuven
Dr. Hester Dibbits	Meertens Instituut
Prof. dr. Martin Everaert	Utrecht Institute of Linguistics OTS
Prof. dr. Joris van Eijnatten	Universiteit Utrecht
Prof. dr. Pol Ghesquière	Katholieke Universiteit Leuven
Drs. Elisabeth d'Halleweyn	Taalunie
Drs. Patricia van Hemert	Vrije Universiteit
Drs. Heidi van den Heuvel	Rijksuniversiteit Groningen
Prof. dr. Frans Hinskens	Meertens Instituut
Drs. Erna Kas	Universiteit Utrecht
Dr. Eloë Kingma	ASCA
Drs. Paul Koopman	Regieorgaan Geesteswetenschappen
Prof. dr. Anne-Marie Korte	Onderzoeksschool NOSTER
Drs. Peter van Laarhoven	Universiteitsbibliotheek Groningen
Dr. Janus Linmans	CWTS
Prof. dr. Eric Moorman	Onderzoeksinstituut voor Historische, Literaire en Culturele Studies
Drs. Hotze Mulder	Universiteit van Amsterdam
Prof. dr. Peter Nissen	NOSTER
Prof. dr. Koen Ottenheim	Onderzoeksschool Kunstgeschiedenis
Prof. dr. Frits van Oostrom	Regieorgaan Geesteswetenschappen
Prof. dr. Maarten Prak	Onderzoeksinstituut voor Geschiedenis en Cultuur
Dr. Frans Ruiters	Onderzoeksinstituut voor Geschiedenis en Cultuur
Dr. Fransisca Snoeck Henkemans	ASCA
Prof. dr. Henriette de Swart	Landelijke Onderzoeksschool Taalwetenschap
Dr. Jeroen Touwen	NW Posthumus Instituut
Prof. dr. Paul Wouters	CWTS
Prof. dr. Jan Luiten van Zanden	Universiteit Utrecht
Prof. dr. Erik-Jan Zürcher	Raad voor de Geesteswetenschappen

Gesprek met het Vlaamse Gezaghebbend Panel voor de Humane en Sociale Wetenschappen, augustus, 2010.

Deelname seminar over prestatie­meting in de geesteswetenschappen, Berlijn januari 2010.

