

KWALITEIT EN RELEVANTIE IN DE GEESTESWETENSCHAPPEN

NAAR EEN ADEQUAAT SYSTEEM VOOR DE BEOORDELING VAN
WETENSCHAPPELIJK ONDERZOEK



2012 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@knaw.nl

www.knaw.nl

pdf beschikbaar op www.knaw.nl

Illustratie omslag: istockphoto

ISBN 978-90-6984-653-8



**KWALITEIT EN RELEVANTIE IN DE
GEESTESWETENSCHAPPEN**

**NAAR EEN ADEQUAAT SYSTEEM VOOR DE
BEOORDELING VAN WETENSCHAPPELIJK ONDERZOEK**

Koninklijke Nederlandse Akademie van Wetenschappen
Eindrapportage van de KNAW Adviescommissie
Kwaliteitsindicatoren Geesteswetenschappen

2012

VOORWOORD

Met het uitbrengen van dit eindrapport van de commissie Kwaliteitsindicatoren voor de Geesteswetenschappen zet de KNAW een volgende stap op weg naar een adequaat systeem voor de beoordeling van wetenschappelijk onderzoek in de geesteswetenschappen. Het bestuur van de KNAW heeft al geruime tijd geleden onderkend dat veel van de indicatoren die worden gebruikt voor de beoordeling van kwaliteit in de natuur- en levenswetenschappen niet als vanzelfsprekend bruikbaar zijn in andere wetenschapsgebieden. Dat geldt voor de geesteswetenschappen, maar ook voor de sociale wetenschappen en de ontwerpende en construerende wetenschappen. Op alle drie deze terreinen heeft de KNAW commissies ingesteld met het verzoek een systeem te ontwerpen dat zowel recht doet aan het werk van onderzoekers in het bewuste wetenschapsgebied als aan het belang van dat gebied voor wetenschap en samenleving. Met het verschijnen van het voorliggende rapport hebben twee van de drie commissies hun advies afgeleverd. Het derde advies – dat van de commissie voor de sociale wetenschappen – wordt eind 2012 verwacht.

Het door de commissie Kwaliteitsindicatoren voor de Geesteswetenschappen gesuggereerde systeem is uitgebreid met het veld besproken en daar bestond veel waardering voor het voorstel. Tezeldertijd is de commissie ook gebleken dat een aantal problemen nog moet worden opgelost voordat het systeem succesvol kan worden toegepast. Het bestuur van de KNAW heeft daarom besloten een tijdelijke stuurgroep in te stellen die de implementatie van de aanbevelingen binnen en door de universiteiten moet bevorderen.

Hans Clevers
President KNAW

INHOUD

SAMENVATTING/SUMMARY 9/11

1. INLEIDING 13
 2. KRITISCHE REFLECTIE OP HET VOORGESTELDE SYSTEEM 17
 - Classificatie van tijdschriften en uitgeverijen 18
 - Verdere uitwerking van de indicatoren 19
 - Indicatoren voor maatschappelijke relevantie 21
 - Ex-postbeoordeling van onderzoeksgroepen en ex-ante-beoordeling van voorstellen 22
 - De *review*-commissie 23
 - Gebruikscontexten 23
 - Verder werken aan de indicatoren 25
 3. DE FOLLOW-UP 29
 - Het eerste spoor: verdere uitwerking van het systeem 30
 4. ADVIES AAN HET BESTUUR KNAW 33
- BIJLAGEN
1. De pilotstudie 35
 2. Opdracht en samenstelling van de commissie 51

SAMENVATTING

Dit is het eindrapport van de KNAW-commissie Kwaliteitsindicatoren voor de geesteswetenschappen, ingesteld in het najaar 2009. Een eindrapport, maar geenszins het einde van een traject. Dit rapport beoogt vooral het gebruik van een nieuw beoordelingssysteem voor onderzoek in de geesteswetenschappen te bevorderen. Het dient in nauwe samenhang te worden gelezen met het interimadvies dat de commissie in 2011 heeft uitgebracht waarin dat systeem wordt voorgesteld. Dat systeem bestaat uit vier elementen: twee beoordelingsaspecten (wetenschappelijke kwaliteit en maatschappelijke relevantie), drie beoordelingscriteria per aspect (output, gebruik en erkenning) en per criterium een aantal indicatoren die veldspecifiek kunnen worden ingevuld. Het vierde element is de *review*-commissie, die op grond van de verzamelde informatie uiteindelijk een oordeel moet vellen. Deze *review*-commissie is in veel gevallen meer dan de traditionele *peer-review*-commissie, er zullen in het algemeen ook externe deskundigen (stakeholders) deel van uitmaken.

Het voorgestelde systeem is in de loop van 2011 uitgebreid in het veld besproken, en de waarde van het systeem is getest in twee pilotstudies bij het Meertens Instituut van de KNAW en het Instituut voor Cultuurwetenschappelijk Onderzoek van de Universiteit van Groningen (ICOG). Er bleek grote waardering voor het voorgestelde systeem, in het bijzonder omdat het recht doet aan de veelzijdigheid van het onderzoek in de geesteswetenschappen, maar tegelijkertijd er in slaagt om relatief eenvoudig en transparant te blijven. De veldraadpleging heeft geen aanleiding gegeven om het voorgestelde systeem in zijn essentie te wijzigen. Wel is duidelijk geworden dat er een aantal problemen is dat nog moet worden opgelost alvorens het mogelijk is dit systeem succesvol toe te passen.

Op grond van het gehele onderzoek komt de commissie met twee aanbevelingen aan het bestuur van de KNAW. De eerste aanbeveling is om het disciplineoverleg letteren en geschiedenis (DLG) te vragen om samen met de twee andere disciplineoverleggen in de geesteswetenschappen (theologie en wijsbegeerte) en de onderzoekscholen te bevorderen dat dit systeem wordt geïmplementeerd in de reguliere evaluatiepraktijk (in de eerste plaats het SEP, maar ook andere waar mogelijk). Prioriteit in dat proces heeft de totstandkoming van een goede tijdschriften-*ranking*.

De tweede aanbeveling is om een onafhankelijke stuurgroep in te stellen die de verdere uitwerking van het systeem bevordert. Deze stuurgroep zou kunnen bestaan uit enkele leden van de huidige commissie (Keimpe Algra en Hans Bennis), de directeur van het Centrum voor Wetenschap en Technologie Studies – CWTS (Paul Wouters), een vertegenwoordiger van de decanen geesteswetenschappen, en de voorzitter van het ERiC-project. De secretaris van de huidige adviescommissie zou ook de secretaris van de stuurgroep moeten zijn.

De taken van deze stuurgroep zijn:

- Het gebruik van dit systeem binnen de geesteswetenschappen in den brede te bevorderen: bij faculteiten, onderzoekscholen, instituten, NWO-aanvragen
- De discussie over deze brede vorm van evaluatie in de adviesraden van de KNAW en breder in het veld, te bevorderen. Dit mede in het licht van het feit dat er nu drie KNAW adviesrapporten zijn (worden) uitgebracht over deze problematiek. Toewerken naar een conferentie of workshop, zo mogelijk met NWO en VSNU, zou daarbij voor de hand liggen.

SUMMARY

This is the final advisory report of the Academy's Committee on Quality Indicators in the Humanities. The Committee was established by the Academy Board in the autumn of 2009 and issued an interim report in June 2011. The present report brings the Committee's work to a close, but it is not the final chapter. The Committee hopes that it represents the first step towards implementing the proposed system, which we see as a suitable method for evaluating humanities research. This report can only be understood in combination with the interim report of 2011, in which the Committee suggested a system consisting of four elements: 1] two evaluation aspects (scientific quality and relevance to society); 2] three assessment criteria for each of these dimensions (output, use and recognition); 3] a number of indicators for each criterion that can be used in a field-specific way; and 4] a review committee to issue a final verdict based on the evidence collected. In many cases, the review committee is likely to be more than a traditional peer review committee. External experts ('stakeholders') should generally be represented on it.

The proposed system was discussed with the field during the pilot phase in 2011. The value and applicability of the system was tested in two pilot studies at the Academy's Meertens Institute and at the Groningen Research Institute for the Study of Culture. Overall, the system met with a lot of a positive response, in particular because it recognises the versatility of research in the humanities while avoiding unnecessary complexities and remaining transparent. The results of the pilot phase did not lead the committee to consider any major changes in the proposed system. However, it became clear that a number of issues need to be addressed before this system can be applied successfully in current evaluation practice. A classification of journals is perhaps the most urgent issue.

Based on the entire project undertaken by the Committee, we are now making two recommendations to the Academy Board.

First, invite the DLG and the deans of theology and philosophy and the research schools to promote the system's implementation in regular evaluation practice (the Standard Evaluation Protocol, but also elsewhere, NWO for example). Priority should be given to establishing a suitable journal ranking.

Second, install a steering committee to elaborate on the system. The steering committee could consist of two members of the current Committee (Keimpe Algra and Hans Bennis), the director of the Leiden Center for Science and Technology Studies - CWTS (Paul Wouters), one of the deans, and the chair of the ERiC project. The secretary of the current committee should also serve as the secretary of the steering committee.

The recommended tasks of the steering committee are:

- To enhance the use of the system in the broad field of humanities research: faculties, research schools, institutes, applications for research funding, etc.
- To promote discussion about this comprehensive form of evaluation in the Academy's advisory councils and elsewhere, with reference to all three advisory reports by the Academy (design and engineering, humanities, and social sciences). A conference or workshop organised with NWO and VSNU would be preferable.

1. INLEIDING

Op verzoek van de Commissie Nationaal Plan Toekomst Geesteswetenschappen ('Commissie Cohen') heeft de KNAW in het najaar van 2009 een adviescommissie ingesteld om een systeem van kwaliteitsbeoordeling te ontwerpen dat aansluit bij de in het geesteswetenschappelijk onderzoek gebruikelijke publicatie- en communicatiepatronen. Net als voor veel andere wetenschapsterreinen (delen van de sociale wetenschappen, ontwerpende en construerende wetenschappen, en delen van de medische en technische gebieden) blijkt de bruikbaarheid van indicatoren gebaseerd op de *citation indexes* voor het merendeel van het geesteswetenschappelijk onderzoek gering.¹ Daarnaast speelt de onderwaardering voor maatschappelijk relevant onderzoek in beoordelingssystemen de geesteswetenschappen parten. De door de KNAW ingestelde Commissie Kwaliteitsindicatoren Geesteswetenschappen (KIG) heeft in 2011 een adviesrapport afgeleverd met daarin een voorstel voor een systeem van evaluatie gebaseerd op criteria en indicatoren die zijn toegesneden op de geesteswetenschappelijke onderzoekspraktijk. Het hier voorliggende rapport dient in samenhang met dat advies te worden gelezen.

In het rapport uit 2011 stelt de commissie KIG dat onderzoekskwaliteit in de geesteswetenschappen niet anders moet worden benaderd dan in andere wetenschapsgebieden. Ook in de geesteswetenschappen bestaat kwaliteit in de betekenis van resultaten van onderzoek voor de wetenschappelijke

¹ De commissie stelt vast dat de bruikbaarheid van citatie-indexen samenhangt met de mate waarin gebieden vertegenwoordigd zijn in de belangrijkste databases, zoals bv. de Web of Science of Scopus, waarop deze indexen zijn gebaseerd. Daarom kunnen ze bruikbaar zijn voor sommige geestes- en sociale wetenschappen (linguïstiek, delen van de geschiedenis, economie, of psychologie) en minder bruikbaar voor sommige exacte vakken.

gemeenschap en daarbuiten. En ook hier zijn het in de eerste plaats *peers*, eventueel aangevuld met andere experts, 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, maar wel om een vrij breed scala van indicatoren voor kwaliteit en relevantie dat recht doet aan de diversiteit aan producten, doelgroepen, en publicatieculturen die men binnen het gebied aantreft.

De commissie heeft een systeem ontworpen 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. Het systeem bestaat uit vier elementen: twee beoordelingsaspecten (wetenschappelijke kwaliteit en maatschappelijke relevantie), drie beoordelingscriteria per aspect (output, gebruik en erkenning) en per criterium een aantal indicatoren die veldspecifiek kunnen worden ingevuld. Het vierde element is de *review*-commissie, die op grond van de verzamelde informatie uiteindelijk een oordeel moet vellen.

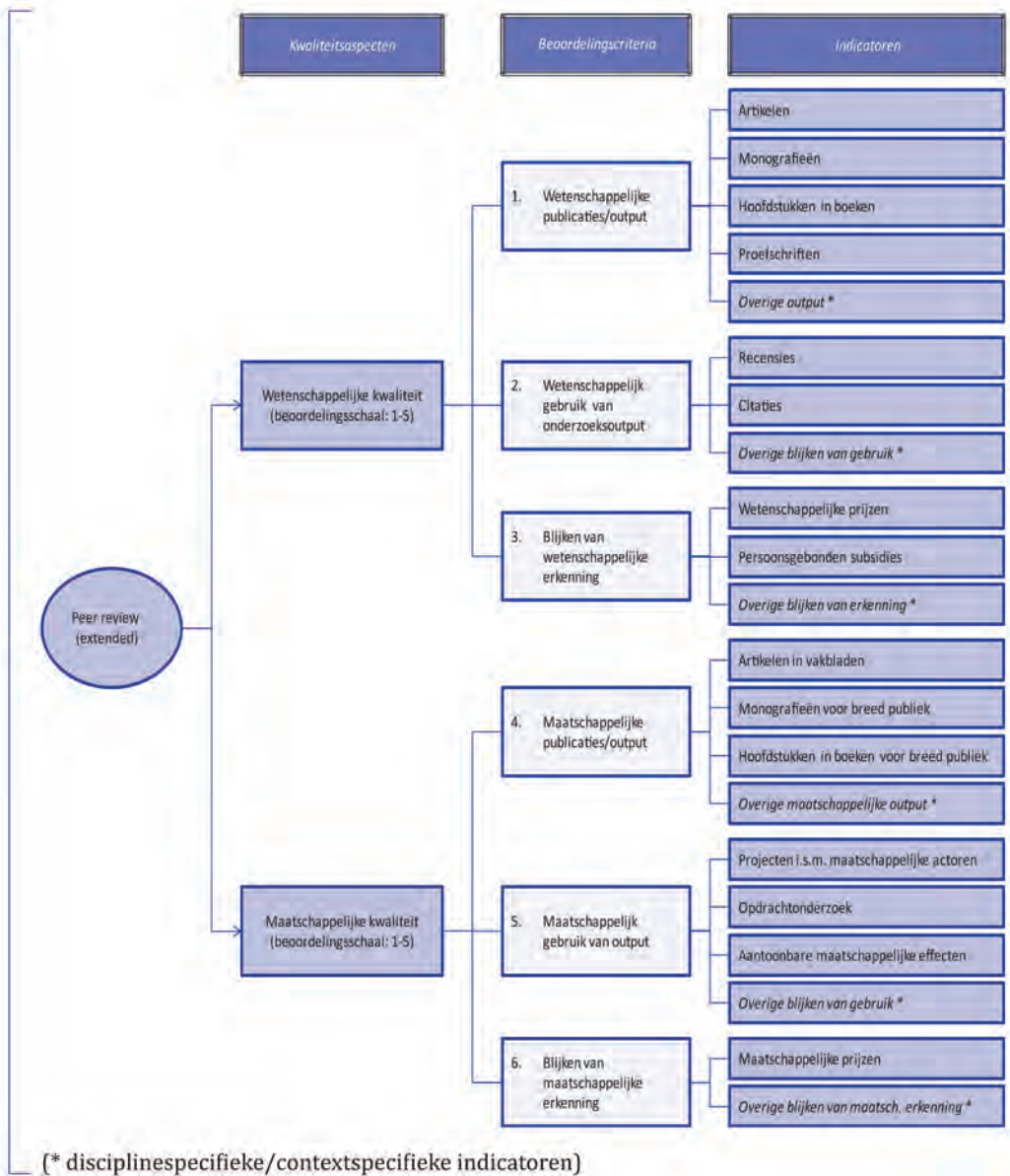
Het systeem is in de loop van 2011 in het veld onderzocht, onder meer in twee *pilots* bij het Meertens Instituut van de KNAW en bij het Instituut voor Cultuurwetenschappelijk Onderzoek (ICOG) van de RUG.

De voornaamste conclusie uit de pilotstudies en de veldraadpleging (zie ook hoofdstuk 2) is dat het door de commissie voorgestelde systeem brede ondersteuning ondervindt in het veld, zowel bij onderzoekers als bestuurders. Er is brede consensus dat de commissie er in geslaagd is een systeem op te zetten dat enerzijds voldoende flexibiliteit kent om de variatie in vakgebieden te representeren en anderzijds de mogelijkheid biedt om eenheden onderling te vergelijken. Er bestaat waardering voor het feit dat er een evenwicht is gevonden voor de beoordeling van de wetenschappelijke kwaliteit en de maatschappelijke relevantie. En het feit dat de weging van de verschillende indicatoren aan de *review*-commissie en/of de opdrachtgever wordt gelaten wordt algemeen gezien als een groot pluspunt.

De commissie heeft tijdens de pilotfase al een aantal stappen gezet om met betrokkenen (decanen, onderzoekscholen, NWO, CWTS) de voortgang van het proces te bevorderen. In het laatste hoofdstuk met de aanbevelingen komen we daarop terug. In het volgende hoofdstuk geven we eerst aan de hand van een aantal thema's de belangrijkste resultaten van de kritische reflectie op het voorgestelde systeem, verkregen via twee pilotstudies en een conferentie. Daarna volgt een opsomming van de voornaamste taken die er liggen om het systeem te vervolmaken en te gebruiken.

Samenvattend: het voorgestelde systeem is goed gevallen binnen de geesteswetenschappelijke gemeenschap; er zijn kritische opmerkingen gemaakt, maar die tasten de essentie van het voorstel geenszins aan. Onze commissie is van mening dat een stevige regie nodig is om het systeem verder te brengen, zowel in bestuurlijke zin als in technisch inhoudelijke zin.

Schema 1: schema van criteria en indicatoren



Toelichting: Het systeem van kwaliteitsindicatoren 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 missie-gerelateerde) 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.

2. KRITISCHE REFLECTIE OP HET VOORGESTELDE SYSTEEM: PILOTSTUDIES EN VELDRAADPLEGING

In dit hoofdstuk presenteren we de voornaamste resultaten van een brede veldraadpleging die geschiedde via verschillende wegen. De werkbaarheid van het systeem van kwaliteitsindicatoren is getoetst in twee pilotstudies bij het Meertens Instituut van de KNAW en bij het Instituut voor Cultuurwetenschappelijk Onderzoek Groningen (ICOG) van de Rijksuniversiteit Groningen. De hoofdvragen in die *pilots* waren: 1. Is het voorgestelde systeem herkenbaar en representatief voor het werk in de geesteswetenschappen; 2. Is het praktisch uitvoerbaar en 3. Wat zijn de belangrijkste vraagstukken die nadere uitwerking behoeven. Verder is op 25 november 2011 een workshop bij het NIAS georganiseerd waarvoor zowel onderzoekers als mensen uit de kringen van beleid werden uitgenodigd. Daar stonden dezelfde drie vragen ter discussie. Ten slotte is er ook nog een focusgroep met beleidsmedewerkers en enkele onderzoekers georganiseerd (26 september 2011). Deze was vooral gericht op een inventarisatie van de praktische problemen die men op de werkvloer voorziet met een dergelijk evaluatiesysteem. In het hier onderstaande ligt de nadruk op de derde vraag, d.w.z. een opsomming van de meest urgente punten die in deze brede veldraadpleging naar voren werden gebracht en besproken, en daarbij telkens de reactie van de commissie. Een meer uitgebreid verslag van de *pilots* en de workshop is te vinden in bijlage 1.

Ten aanzien van de eerste twee vragen kunnen we hier kort zijn. Het belangrijkste is te vermelden dat er brede instemming is gebleken met de door de commissie voorgestelde aanpak. Zowel van de zijde van de decanen als de onderzoekscholen is het voorgestelde systeem geprezen om zijn vermogen aan de diversiteit van de geesteswetenschappelijke output recht te doen en tegelijkertijd voldoende systematiek te herbergen om tot op zekere hoogte vergelijkingen mogelijk te maken. Men vindt het systeem een duidelijke verbetering ten opzichte van de bestaande praktijk. De gelijk-

waardige plaats voor boeken en hoofdstukken in boeken wordt zeer gewaardeerd, evenals het uitgangspunt dat uit het systeem spreekt dat geesteswetenschappen worden gekenmerkt door een grote diversiteit aan producten en activiteiten en dikwijls door een missie die zowel gericht is op de wetenschappelijke als de maatschappelijke betekenis van het onderzoek.

Ook het gegeven dat het uiteindelijke oordeel ligt bij een commissie van deskundigen (*peers*, of *extended peers*) ondervindt veel instemming. Het voorgestelde systeem heeft als uitgangspunt dat er niet een a priori verschil in gewicht is tussen beide hoofdcriteria, wetenschappelijke kwaliteit en maatschappelijke relevantie. Het laat de afweging over aan de *review*-commissie (in overleg met de opdrachtgever) omdat men het vertrouwen heeft dat door die experts de juiste afweging kan worden gemaakt. Er bestond wel enige twijfel of dat ook inderdaad gebeurt, vooral omdat de informatie over wetenschappelijke kwaliteit over het algemeen robuuster is dan die over de maatschappelijke relevantie. Algemeen is dan ook de gedachte dat er vooral aan de indicatoren voor maatschappelijke relevantie nog het nodige ontwikkelwerk moet worden gedaan. Het systeem lijkt bovendien zeer uitvoerbaar, en het wordt gezien als een duidelijke handleiding voor het schrijven van de zelfevaluaties die nodig zijn voor het SEP. Er bestaat wel enige zorg over de beschikbaarheid van voldoende robuuste data - een deel van de data die voor dit systeem nodig zijn worden eenvoudigweg nog niet verzameld door de instellingen. Niettegenstaande de positieve ontvangst van het voorgestelde systeem, bleek tijdens de veldraadpleging ook dat er nog tal van onderwerpen zijn die nadere uitwerking vragen. In het onderstaande behandelen we de meest urgente van deze punten.

Classificatie van tijdschriften en uitgeverijen

De noodzaak van een goede tijdschriftenclassificatie wordt door velen genoemd. Er lopen verschillende projecten, nationaal en internationaal, waarbij wordt getracht tot een classificatie van tijdschriften te komen. In Noorwegen werkt men met een onderscheid tussen een basisniveau (meest nationale tijdschriften) en een selectniveau (meest internationaal), in België is er een zogenoemd gezaghebbend panel dat de classificatie voor verschillende disciplines in de sociale en humane wetenschappen coördineert. De ESF heeft een poging gedaan om een Europese classificatie te maken, maar dat is nog niet gelukt. Lokaal en disciplinair wordt er in Nederland aan diverse lijsten gewerkt. De decanen hebben aangegeven dat zij, in overleg met de onderzoekscholen, wel tot een vaststelling van een lijst kunnen komen. Deze is misschien niet honderd procent perfect, maar er kan wel voldoende overeenstemming worden bereikt over de overgrote meerderheid van de tijdschriften. Er bestaat wel enige reserve bij sommigen. Gewaarschuwd wordt er voor het gevaar van inperking: tijdschriften die van belang zijn voor de (nieuwe) ontwikkelingen in een vakgebied of tussen vakgebieden kunnen in het gedrang komen omdat bij de samenstelling van lijsten de gevestigde orde onder de tijdschriften dominant zou zijn. Ook wordt er aandacht gevraagd voor

het belang van Nederlandstalige tijdschriften voor sommige vakgebieden. Voorkomen zou moeten worden dat er een classificatie wordt ingevoerd waarbij deze tijdschriften niet voor de hoogste categorie in aanmerking komen (zoals in het Noorse systeem het geval lijkt).

De commissie meent dat er in het algemeen aandacht moet zijn voor de gevaren van het gebruik van tijdschriftenlijsten op lokaal, of facultair niveau, maar zij acht het niet haar taak hierover uitspraken te doen. Volgens de commissie ligt dat op de weg van gezaghebbende organen in het vakgebied. Voorts wijst de commissie er op dat een groter kader dat faculteiten overstijgt kan voorkomen dat de lijsten al te particulier worden ingevuld. Het is raadzaam om de lijsten die er nu al functioneren te verzamelen en te bezien wat voor effecten die lijsten hebben. Een van de uitkomsten van de pilotfase is dat er door Hotze Mulder, hoofd afdeling onderzoek, Bureau Faculteit der Geesteswetenschappen UvA, het initiatief is genomen tot zo'n inventarisatie².

De commissie is van mening dat de tijd rijp is om snel tot een vaststelling van een procedure te komen om tot een dergelijke tijdschriftenlijst te komen, dit in overleg met betrokken partijen: decanen, onderzoekscholen, KNAW (eventueel VSNU en NWO). Datzelfde geldt *mutatis mutandis* voor het vaststellen van een onderscheidend criterium voor uitgeverijen. Dat ligt onder meer ingewikkeld omdat het criterium *peer reviewed* niet eenduidig kan worden gehanteerd. Uitgeverijen hanteren verschillende vormen van *peer review*, en uitgeverijen die geen *peer review* hanteren kunnen toch kwalitatief goede uitgaven hebben.

Verdere uitwerking van de indicatoren

De rol van de indicatoren in het systeem

Het model dat de commissie voorstelt spreekt aan: de *review*-commissie geeft een eindoordeel op basis van een afweging van zes criteria die de breedte van geesteswetenschappelijk onderzoek dekken. De indicatoren hebben primair tot doel systematische informatie te leveren voor die criteria. Met andere woorden, het conceptuele onderscheid tussen indicatoren en criteria is essentieel voor de flexibiliteit van het systeem. Alle (sub)gebieden worden beoordeeld op dezelfde zes criteria, en kunnen op grond daarvan worden vergeleken. Verschillen tussen gebieden kunnen tot uitdrukking worden gebracht in de keuze voor een gebiedspecifieke set indicatoren die de criteria onderbouwt. Die keuze vergt wel een goede structuur voor de wijze waarop het

² Een eerste inventarisatie toont aan dat weliswaar al veel is gedaan om te komen tot een classificatie van tijdschriften en andere output-media in de geesteswetenschappen, maar dat er van consensus nog geen sprake is. Het voorstel is om eerst overeenstemming te bereiken (op basis van wat nu al beschikbaar is) over criteria voor drie categorieën publicaties, A, B en C. Denk bv. aan criteria als de mate van onafhankelijke *peer review*; afwijzingsgraad; internationaliteit (evt. gekoppeld aan taal: vaak worden Frans, Duits, Engels genoemd als de internationale talen, maar voor klassieke archeologie zou bijvoorbeeld Italiaans ook als internationaal kunnen gelden); verzorging van publicatie; verspreidingsgraad; kwaliteitstraditie (hoe lang al op dit niveau). De decanen hebben aangegeven bij de ontwikkeling van zo'n classificatie het voortouw te willen nemen.

veld betrokken is bij de vaststelling van de indicatoren. Naar de mening van de commissie moet daartoe een relevant aggregatieniveau worden gekozen, bij voorkeur niet te kleine gebieden, maar vakgebied overschrijdende domeinen. Deze zouden landelijk moeten worden vastgesteld, in overleg met decanen en onderzoekscholen. Vervolgens moeten in die domeinen keuzes worden gemaakt.

Ten aanzien van de indicatoren die worden voorgesteld is opgemerkt dat het eigenlijk veeleer categorieën zijn voor indicatoren. De commissie is het daarmee eens, maar geeft daarbij aan dat zij het precies als haar taak zag om categorieën van indicatoren te ontwikkelen en dat een verdere verfijning van het systeem zelf (anders dan in de toepassing van het systeem door gebruikers) tot een onhanteerbare omvang leidt. Voor de verdere ontwikkeling van indicatoren is het veld aan zet, samen met experts op het gebied van indicatoren en dataverzameling. Die samenwerking van experts en domeinen is nodig om te voorkomen dat het een technocratische operatie wordt.

Vaste basisset of niet

Het voorstel van de commissie is om wel een totale set van indicatoren te hanteren (de laatste kolom in figuur 1), maar de keuze voor welke indicatoren in een domein worden gebruikt over te laten aan betrokkenen in dat domein. Anderen spreken zich uit voor een set centrale (landelijke, gelijk voor alle faculteiten) indicatoren. Daarmee zou onderlinge vergelijking beter mogelijk zijn. De flexibiliteit blijft in die visie aanwezig dankzij de indicator categorie 'overig'. Het punt van vergelijkbaarheid is belangrijk voor bv. NWO omdat in een aantal gevallen over alle disciplines heen wordt beoordeeld: bij VICI bijvoorbeeld moet een afweging gemaakt worden tussen aanvragen van alfa, bèta, gamma en medisch.

De commissie stelt zich op het standpunt dat ook in haar uitleg van het systeem de vergelijkbaarheid aanwezig is, omdat namelijk alle gebieden op de zes criteria kunnen worden vergeleken. Vergelijking tussen vakgebieden op basis van een vaste set indicatoren blijft lastig vanwege verschillende tradities, groepspublicaties versus solo bijvoorbeeld. De commissie verwijst ook naar het CWTS-onderzoek dat kijkt naar verschillen in publicatiecultuur, bij rechten en geschiedenis aan de EUR. Dat onderzoek loopt min of meer parallel aan het KNAW-onderzoek en de voorlopige resultaten daarvan bevestigen het probleem van die verschillen voor vergelijkende evaluatie. De onderzoekers concluderen dat de analyse van de output en impact niet op dezelfde manier kan worden uitgevoerd voor verschillende vakgebieden. Dat komt deels door de diversiteit in publicatietypen en -media, maar achterliggend mogelijk ook door verschillen in publicatiestrategie van onderzoekers. Tegelijkertijd concluderen de onderzoekers dat door alle vakgebieden in tijdschriften wordt gepubliceerd, maar dat sommige vakgebieden wel en andere niet in Web- of Science-tijdschriften kunnen publiceren. En het is juist die database die het meest wordt gebruikt voor vergelijkend bibliometrisch onderzoek.

Nieuwe ontwikkelingen

Van verschillende kanten is er op gewezen dat er naast verschillen in publicatiecultuur tussen vakgebieden, ook andere verschillen bestaan die tijds- en generatiegebonden zijn. Jongere onderzoekers doen steeds meer digitaal, maken gebruik van Facebook en Twitter. Digitalisering heeft een vergaande invloed op het systeem van *scholarly communication*. Wetenschappers communiceren op verschillende manieren met elkaar binnen en over de grenzen van hun vakgebied. Gedrukte publicaties, schriftelijke correspondentie en de schriftelijke neerslag van conferenties, zijn aangevuld met e-mail, *science blogs*, *e-books*, *e-journals* en zelfs *digital laboratories*. Gaandeweg verplaatst ook de wetenschappelijke uitgeverij zich naar het digitale domein. In het domein van de monografieën ligt een online toekomst in het verschiet, zoals dat al het geval is bij de wetenschappelijke tijdschriften. Ook de ontwikkelingen rond *open access* worden in dit verband genoemd. Lezen via *open access* is inmiddels een veelvoud van de uitleen van boeken. De vraag is hoe publiceren in dergelijke media moet worden meegewogen. Behalve voor de innovatie van het uitgeefplatform (tijdschrift en boek), heeft digitalisering ook belangrijke gevolgen voor het veld van wetenschap zelf, op basis van innovaties (1) in het proces van onderzoek en wetenschap, (2) in de organisatie van de wetenschappelijke communicatie en (3) in de politiek-economische constellatie van het wetenschappelijk uitgeven.

De commissie realiseert zich dat daarbij oude vragen in een nieuw daglicht komen te staan: wat is de rol van peer review in het digitale tijdperk, op welk moment is een publicatie ook een publicatie (is een blog een publicatie?). Digitalisering leidt tot meer dynamische productie, interactiever (dus ook minder goed toe te kennen aan individuen of groep), vluchtiger ook. De indicatoren zijn nu veelal gericht op 'stollingsmomenten' maar dat houdt geen rekening met de nieuwe dynamiek in de wetenschappelijke gemeenschap.

Bij de verdere ontwikkeling van het systeem is speciale aandacht nodig voor bijzondere categorieën van output, zoals bv. databases, films, audio. Daarnaast is het ook zo dat communicatie over onderzoek, zeker bij de jongere generatie steeds meer via sociale media geschiedt. Er zou daarbij ook aandacht moeten zijn voor nieuwe vormen van kwaliteitscontrole en de veranderende rol van *peer review*.

Concluderend is de commissie van mening dat deze ontwikkelingen zeker de aandacht behoeven in het vervolgproces, maar zij ziet op dit moment nog onvoldoende aanleiding het ontworpen systeem te wijzigen.

Indicatoren voor maatschappelijke relevantie

Ten aanzien van de maatschappelijke relevantie worden drie soorten van opmerkingen gemaakt. In de eerste plaats bestaat er algemene overeenstemming dat dit aspect in belang niet onderdoet voor de wetenschappelijke kwaliteit. Beide aspecten verdienen een gelijkwaardige en evenwichtige afweging in de beoordeling van gees-

teswetenschappelijk onderzoek. In hoeverre dat mogelijk is, en dat is het tweede punt, is echter een punt van discussie. De ontwikkeling van indicatoren voor maatschappelijke relevantie van onderzoek is minder ver dan die voor de wetenschappelijke kwaliteit. Wat de ontwikkeling van indicatoren voor de maatschappelijke kwaliteit betreft wordt in de *pilots* vaak verwezen naar het ERiC-project en naar de Nationale Commissie Valorisation. Het ERiC-project heeft een raamwerk ontwikkeld dat gelijk is voor alle contexten, maar daarbinnen is een grote vrijheid indicatoren te gebruiken. De Nationale Commissie Valorisation zit op een andere lijn waarbij een systeem wordt ontwikkeld dat uitgaat van een vaste lijst met indicatoren die voor elk vakgebied zouden moeten gelden. In het vervolgotraject zou naar deze beide benaderingen moeten worden gekeken, waarbij de commissie aantekent dat de contextuele benadering van het ERiC-project meer overeenkomsten vertoont met de benadering van het systeem dat zij voor de geesteswetenschappen in gedachten heeft. Ten slotte, het derde punt, maken deelnemers aan zowel de conferentie als de focusgroepen zich zorgen over de vraag hoe de afweging tussen beide aspecten plaats vindt in de *review*-commissie. Immers, er is daar weinig ervaring mee, zowel wat betreft het werken met gemengde commissies, als met de afweging van (meestal meer robuuste) indicatoren voor de wetenschappelijke kwaliteit en indicatoren voor maatschappelijke relevantie. Naar het oordeel van onze commissie is er inmiddels wel een behoorlijke aandacht voor deze vragen, zeker in beide genoemde projecten, maar zal er ook sprake zijn van een al doende leren. De commissie tekent daarbij aan dat men zich niet blind moet staren op de exclusieve robuustheid van de kwantitatieve indicatoren zoals die veelal ten aanzien van de wetenschappelijke kwaliteit worden gebruikt. Ook kwalitatieve indicatoren (vaak gebruikt bij het identificeren van maatschappelijke relevantie) kunnen robuuste informatie geven. Overigens zijn er, mede dankzij het internet, steeds meer mogelijkheden om ook de maatschappelijke output van onderzoeksgroepen te traceren en meer kwantitatieve benaderingen te hanteren.

Ex-postbeoordeling van onderzoeksgroepen en ex-ante-beoordeling van voorstellen.

Van verschillende kanten wordt er op gewezen dat het systeem vooral gericht lijkt op de beoordeling achteraf van groepen of vergelijkbare eenheden, en veel minder geschikt is voor exantebeoordeling van voorstellen, bijvoorbeeld bij NWO of in de Europese programma's. De commissie geeft aan dat het systeem inderdaad in de eerste plaats is gericht op de praktijk van de externe beoordelingen volgens het SEP. Daarbij wordt vooral teruggekeken naar de prestaties in het recente verleden (ofschoon in het SEP ook nadrukkelijk wordt gesteld dat toekomstige perspectieven in beschouwing moeten worden genomen). Het systeem is naar het oordeel van de commissie bruikbaar bij ex-post-beoordelingen in het algemeen en bij het ex-post-gedeelte van exantebeoordelingen (bij het beoordelen van een cv van een projectaanvrager, bijvoorbeeld). Voor het pure ex-antedeel van een beoordeling (de te verwachten kwaliteit, opbrengst

en haalbaarheid van een onderzoeksvorstel) zijn eigenlijk geen indicatoren te formuleren. Er zijn wel criteria die bv. door NWO worden gehanteerd, zoals helderheid, gestructureerdheid, innovativiteit; maar de vraag of die worden gehaald is alleen te beantwoorden via een inschatting van de *reviewers*.

De review-commissie

Hoewel de pilotfase zich grotendeels concentreerde op het systeem en de uitvoerbaarheid ervan, werd er ook wel aandacht besteed aan het functioneren van de *review*-commissie, in het bijzonder voor wat betreft de samenstelling en de weging van de indicatoren. De kritiek was dat er voor de leden van een *review*-commissie een zware rol is weggelegd, zij moeten immers de afweging maken op grond van indicatoren en criteria die heel verschillend onderbouwd zijn, en in veel gevallen zijn deze commissies bovendien gemengd van samenstelling. Er is met dit soort processen nog niet heel veel ervaring, en afgezien van de onzekerheden in de groepsdynamiek, zijn er vragen bij het traject daaraan voorafgaand. Hoe worden die commissies samengesteld, en door wie, en wat voor instructies krijgen zij mee? Als het meer is dan traditionele *peer review*, hoe wordt dan bepaald wie de andere deelnemers zijn. En wat zijn andere manieren om externe expertise in het *review*-proces te gebruiken? Al deze vragen en meer zijn niet uitgewerkt, terwijl het toch van cruciaal belang lijkt voor het functioneren van dit systeem. Van verschillende kanten werd gesuggereerd dat het systeem daarom zo snel mogelijk als een experiment moest worden gestart om te leren van de praktijk. De commissie erkent dat er over de rol en het functioneren van de *review*-commissie meer moet worden nagedacht. De commissie juicht het toe als er in de komende tijd veel met het systeem wordt geëxperimenteerd, en dat kan ook in de aanloop naar het nieuwe SEP in 2015.

Gebruiksccontexten

In de discussie over de bruikbaarheid van het systeem in verschillende contexten kwamen diverse kwesties naar voren. Algemeen is de verwachting dat het systeem een beter beeld kan geven in het kader van externe (SEP-)onderzoeksbeoordelingen, maar ook ziet men dat het systeem mogelijkheden biedt in het kader van het aanstellingsbeleid, de begeleiding van promovendi en het personeelsbeleid (persoonlijke voortgangsgesprekken). Opgemerkt werd dat het voor jongere onderzoekers van belang is dat er een duidelijker kader van verwachtingen komt dan voorheen, met een nationaal en bij voorkeur ook internationaal perspectief. Bij het verder uit te werken systeem zou ook met die contexten rekening moeten worden gehouden.

De bruikbaarheid van het systeem is voorts gebaat bij aandacht voor de praktische kant van de dataleverantie. Gegevens over publicaties dienen eenvoudig aan te leveren en te verwerken te zijn. Dat ligt voor de hand, maar is niet eenvoudig omdat er sprake

zal zijn van een grotere diversiteit aan te registreren onderzoeksresultaten. Een ander probleem is dat de aanlevering van de huidige meer beperkte informatiestroom al onder druk staat omdat onderzoekers het belang ervan onvoldoende onderkennen. Dit laatste wordt enerzijds in de hand gewerkt doordat in de universitaire beloningssystemen slechts een beperkt aantal indicatoren belangrijk worden geacht, en anderzijds omdat bepaalde categorieën van onderzoek (nog) niet aangemeld kunnen worden. Bij beleidspersonen en bij onderzoekers leeft de verwachting dat een overkoepelend representatief systeem richtinggevend kan zijn voor de ontwikkeling in de gewoontes met betrekking tot publiceren.

Beschikbaarheid van informatie

De informatievoorziening voor de indicatoren bleek een belangrijk discussiepunt. Welke informatie is al beschikbaar en welke (technische) mogelijkheden zijn er om de in het systeem voorgestelde nieuwe informatie te verkrijgen? Het huidige Research Informatie Systeem METIS blijkt meerdere versies te hebben zodat er op dit moment nog geen sprake is van een voor alle instellingen eenduidig stelsel: het Meertens Instituut blijkt bijvoorbeeld een grotere reikwijdte aan publicaties in METIS onder te kunnen brengen dan het ICOG.

METIS biedt weliswaar ruimte aan diverse vormen van publicaties, zoals boeken (wetenschappelijke monografieën) en bijdrage in bundels (hoofdstukken in boeken), en artikelen in wetenschappelijke tijdschriften. En ook is er ruimte om informatie te verwerken met betrekking tot maatschappelijke publicaties, zoals artikelen in vakbladen, en monografieën voor breed publiek en dergelijke. Maar de registratie van deze publicaties wordt bemoeilijkt doordat bij de invoervelden de software van METIS vraagt naar extra informatie, ter validatie van de in te voeren publicatiekenmerken. Dit leidt in een onbekend maar vermoedelijk substantieel aantal gevallen tot een *mismatch*. De gevraagde informatie bij de datavalidatie sluit niet altijd aan bij kenmerken van de publicaties zoals die gebruikelijk zijn in de geesteswetenschappen. Het gevolg is dat de gegevens van monografieën, wetenschappelijke artikelen en dergelijke worden gerangschikt onder 'overige publicaties', wat leidt tot een onderschatting van het aantal publicaties in de relevante categorieën.

Gegevens over wetenschappelijk gebruik en blijken van wetenschappelijke erkenning ontbreken veelal of worden niet centraal verzameld. Zo zal wel bekend zijn welke wetenschappelijke prijzen en persoonsgebonden subsidies er verkregen zijn, maar het is nu niet mogelijk de informatie op te nemen in de huidige onderzoeksinformatiesystemen. Ook de gegevensverzameling van indicatoren voor het maatschappelijk effect of gebruik ontbreekt nog grotendeels, evenals de gegevens over de blijken van maatschappelijke erkenning. Ook hier geldt dat de informatie deels al wel bekend is, bij onderzoekers of wellicht ook bij communicatieafdelingen van faculteiten of universiteiten.

De commissie is van mening dat invoering van data in METIS op zichzelf niet een groot probleem is, er zijn immers 'vrije' plaatsen beschikbaar voor nieuwe informa-

tie. Maar de commissie ziet ook dat er in de praktijk diverse problemen zijn die de invoering bemoeilijken. Daarbij is het in ieder geval van belang oplossingen te kiezen die landelijk per domein worden gedragen. Motivatie om gegevens aan te leveren is een cruciale succesfactor. Hoewel sommige informatievelden al wel beschikbaar zijn wordt de benodigde informatie kennelijk niet of slechts beperkt aangeleverd. Met het oog daarop acht de commissie het van belang om voor medewerkers een duidelijke waarderings- of beloningscontext te bieden waarin het voor hen van belang is om ook die informatie aan te bieden.

Verder werken aan de indicatoren

De voorgestelde indicatoren worden over het geheel genomen herkend en gewaardeerd. In feite is in het systeem sprake van indicatorcategorieën, dat wil zeggen bij de verdere uitwerking van het systeem zullen per categorie beslissingen moeten worden genomen over welke data moeten worden verzameld. Hieronder zullen we per indicatorcategorie aangeven welke stappen er nog moeten worden genomen om tot een werkbaar systeem te komen. Maar eerst nog enkele opmerkingen die in het kader van de discussie over indicatoren werden gemaakt en die bij de verdere ontwikkeling overwogen kunnen worden.

Veel belang werd gehecht aan de verhouding publicaties in internationale tijdschriften, eventueel als een aparte indicator, versus Nederlandstalige publicaties. Voor sommige gebieden is het vanzelfsprekend om in het Nederlands te publiceren. Deze kwestie hangt nauw samen met de tijdschriftenkwalificatie. De commissie neemt hierover geen standpunt in omdat zij dit nu juist de verantwoordelijkheid van het vakgebied vindt.

Aandacht werd gevraagd voor de nieuwere vormen van wetenschappelijke en / of maatschappelijke output, zoals (*gereviewde*) catalogi bij tentoonstellingen, dvd-producties of internetproducten (zoals al dan niet publiekstoegankelijke databases). Enerzijds merken deelnemers op dat het ontbreken van dergelijke categorieën ertoe kan leiden dat onderzoekers deze vormen van output negeren, ook al is dat voor de uitoefening van hun vak relevant. Dit is bijvoorbeeld het geval met boekbesprekingen die niet geregistreerd worden als wetenschappelijke output. Anderzijds blijken er in METIS wel degelijk categorieën te zijn, zoals bijvoorbeeld dvd-producties, die door deelnemers niet worden benut omdat deze binnen de context van hun instituut geen relevant prestatie criterium zijn. Het door de commissie voorgestelde systeem voorziet weliswaar in de mogelijkheid deze nieuwe categorieën op te nemen, maar hier speelt het al eerder genoemde afstemmingsprobleem op het METIS-systeem.

Voorts werd opgemerkt dat indicatoren voor erkenning moeilijk te verzamelen blijken. Deelnemers vinden dergelijke indicatoren wel van belang, niet alleen voor gevestigde onderzoekers, maar ook in situaties waarin een onderzoeker, zoals een promovendus, nog onvoldoende *track record* heeft opgebouwd. Maar met name in

dat laatste geval werkt simpelweg tellen niet, dan zal veel meer kwalitatief moeten worden gekeken. Het verkrijgen van beurzen, prijzen en het verwerven van opdrachtinkomsten zijn daarbij voor de hand liggende indicatoren.

Ten slotte werd er nogal veel gediscussieerd over het gewicht van de verschillende indicatoren. De weging van de verschillende indicatoren wordt in het voorstel van de commissie overgelaten aan de *reviewers* en/of de opdrachtgever. Om een goede weging van de verschillende indicatoren te vergemakkelijken is het wel noodzakelijk dat de indicatoren zo helder en robuust mogelijk zijn. Bij een aantal categorieën is het mogelijk om gewichten toe te kennen aan bepaalde indicatoren, bijvoorbeeld als er overeenstemming bestaat over een tijdschriftenclassificatie. Voor uitgeverijen zal dit een stuk lastiger zijn omdat zich tussen en ook binnen uitgeverijen grote variaties kunnen voordoen in fondsen of reeksen en de daarbij behorende vormen van peer review. Dat maakt het voor externe beoordelaars lastig om tot een afgewogen oordeel te komen over het belang van verschillende publicaties. De commissie stelt zich ook hier op het standpunt dat er ten aanzien van dit soort problemen via landelijke coördinatie een classificatie tot stand moet komen.

Tot slot van dit hoofdstuk geven wij per indicator aan wat de meest urgente kwestie is bij de verdere ontwikkeling.

Tabel 1 voorzien traject bij de uitwerking van het systeem naar indicatoren

Beoordelingscriterium	Indicatorcategorie	Informatie/data problematiek
1. Wetenschappelijke publicaties	Artikelen	Lijst, eventueel aangevuld met sleutelpublicaties; primair zal een classificatie van tijdschriften moeten worden ontwikkeld
	Monografieën	Lijst, eventueel aangevuld met sleutelpublicaties; primair zal een classificatie van uitgevers moeten worden ontwikkeld
	Hoofdstukken in boeken	Lijst, eventueel aangevuld met sleutelpublicaties; primair zal een classificatie van uitgevers moeten worden ontwikkeld
	Afgeronde proefschriften	- Lijst per onderzoeksgroep - Lijst per onderzoeker (promotor)
	<i>Overige wetenschappelijke output</i>	Het veld beslist welke indicatoren van belang zijn, voorbeelden: boekrecensies, databases, redacties, handboeken.
Wetenschappelijk gebruik van onderzoeksoutput	Recensies	Naspeurbare verwijzingen naar samples
	Bibliometrische indicatoren	Citatieanalyse (mits de databases voldoende dekkend zijn en aansluiten bij de diversiteit van publicatieculturen)
	<i>Overige blijken van gebruik</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</i>	Het veld beslist welke indicatoren van belang zijn, voorbeelden: fellowships bij prestigieuze (buitenlandse) instituten, chief/full editorships van gerenommeerde (inter)nationale tijdschriften/boeken/bundels, adviseurs-/referentieschappen (bijv. NWO, visitatiecommissies) etc.
4. Maatschappelijke publicaties	Artikelen in vakbladen (niet primair wetenschappelijke tijdschriften)	- Lijst, eventueel aangevuld met sleutelpublicaties - internetsearchtechnieken
	Monografieën voor niet-wetenschappelijke professionals en geïnteresseerden	- Lijst, eventueel aangevuld met sleutelpublicaties internetsearchtechnieken
	Hoofdstukken in boeken voor niet-wetenschappelijke professionals en geïnteresseerden	- Lijst, eventueel aangevuld met sleutelpublicaties - internetsearchtechnieken
	<i>Overige maatschappelijke output</i>	Het veld beslist welke indicatoren van belang zijn, voorbeelden: bundels voor niet-wetenschappelijke professionals en geïnteresseerden, redacties van vakbladen, handboeken, woordenboeken, tekstedities, databases, software, tentoonstellingen, catalogi, vertalingen, beleidsadviesrapporten
5. Maatschappelijk gebruik van onderzoeksoutput	Projecten uitgevoerd in samenwerking met maatschappelijke actoren	Vermelding met jaartal
	Opdrachtonderzoek	Vermelding met jaartal
	Aantoonbare maatschappelijke effecten van onderzoek	Vermelding met jaartal
	<i>Overige vormen van maatschappelijk gebruik</i>	Het veld beslist welke indicatoren van belang zijn, ook in relatie met andere maatschappelijke outputindicatoren, bijv. recensies, citaties in (beleids)rapporten, gebruik van publicaties, media-aandacht, verkochte/uitgeleende boeken
6. Blijken van maatschappelijke erkenning	Maatschappelijke prijzen	Vermelding met jaartal
	<i>Overige blijken van maatschappelijke erkenning</i>	Per context nader te bepalen welke kwantitatieve en/of kwalitatieve informatie verlangd wordt. bijv. maatschappelijke benoemingen, lezingen, mediaoptredens, adviesfuncties

In het vervolgtraject met betrekking tot de indicatoren kan worden samengewerkt met het CWTS dat momenteel een studie uitvoert bij de Erasmus Universiteit voor de vakgebieden geschiedenis en rechtsgeleerdheid. De onderzoekers hebben op grond van testen met METIS-gegevens geconstateerd dat de zichtbaarheid van de geesteswetenschappen in het *Web of Science* (WoS) veel lager is dan die van het economisch of biomedisch onderzoek van de EUR. Een niet onverwachte uitkomst; de conclusie is dan ook dat het niet veel zin heeft om de onderzoeksactiviteiten van de geesteswetenschappen in de WoS te traceren voor bibliometrische indicatoren. In het onderzoek werd ook duidelijk dat de niet biomedische onderdelen van de EUR veel andere typen van wetenschappelijke communicatie registreren, zoals boeken, boekhoofdstukken, conferentiebijdragen, annotaties (in het geval van de rechtsgeleerdheid), rapporten, magazinepublicaties, en een categorie overige typen, die o.a. informatie bevat over ontvangen prijzen, deelname aan *review*-commissies, redacties, etc.).

3. DE FOLLOW-UP

De commissie trekt in dit hoofdstuk conclusies uit de discussies met het veld voor het vervolgtraject. Er zijn in de ogen van de commissie twee sporen voor de follow-up. In de eerste plaats moet er worden gewerkt aan de verdere uitwerking van het voorgestelde systeem, en in de tweede plaats zal er een beslissing moeten worden genomen over het bestuurlijke proces dat nodig is om het systeem te implementeren in de context van het SEP en eventueel andere contexten (NWO, EU).

Hieronder zullen we kort ingaan op een aantal kwesties die zullen moeten worden aangepakt in het traject naar een nieuw beoordelingssysteem voor geesteswetenschappelijk onderzoek. Uit de opsomming blijkt dat er nog heel wat werk moet worden verzet en het dus noodzakelijk is om goed na te denken over de regie van dit traject. In principe kan voor dit traject de tijd worden genomen indien men het nieuwe SEP in 2015 als richtdatum kiest. Tot die tijd is het dan mogelijk binnen bestaande evaluatiecontexten te experimenteren met dit systeem en het zodoende verder te ontwikkelen. Dit is in overeenstemming met afspraken die vorig jaar zijn gemaakt in het rectorenoverleg van juni, in aanwezigheid van de president van de KNAW.³

³. De rectoren, in het bijzijn van de president van de KNAW, waren destijds van oordeel dat de universiteiten/faculteiten komende jaren vooral door moeten gaan met ervaring op te doen met de adviezen van TWINS, geesteswetenschappen, ERiC, e.d. Zij zullen deze ervaringen blijven uitwisselen. Een aanpassing van het SEP voor 2013 wanneer een *mid-termreview* is gepland, werd niet opportuun geacht. (verslag rectorenoverleg 8 juni 2011)

Het eerste spoor: verdere uitwerking van het systeem

Tijdschriftenkwalificatie

Er is grote consensus dat een goede tijdschriftenclassificatie een belangrijke voorwaarde is voor het goed functioneren van het systeem. Er is tevens een groeiende overeenstemming in het veld dat dit met niet al te veel moeite bereikbaar is (decanen-overleg dd. 11 november 2011).

De decanen geesteswetenschappen hebben aangegeven dat zij hierbij een leidende rol willen spelen. Naar aanleiding van de bijeenkomst met beleidsmedewerkers in september 2011 is door Hotze Mulder (UvA) een eerste aanzet gegeven voor het tot stand brengen van een dergelijke lijst. Voorgesteld wordt eerst te kijken naar zowel internationale voorbeelden als naar lokale initiatieven en op grond van deze inventarisatie een besluit te nemen.

Voor de uitwerking van tijdschriftenlijsten zou het overleg van de decanen letteren en geschiedenis (DLG) moeten samenwerken met LOGOS en de andere discipline overleg organen om de lijsten te gaan samenstellen aan de hand van de inventarisatie.

Uitgeverijen

Het classificeren van uitgevers is veel lastiger. Eerder werd al aangegeven dat het gebruik van *peer review* niet altijd een onderscheidend criterium is, nog afgezien van het feit dat *peer review* op verschillende manieren wordt gebruikt en ook niet altijd voor het gehele fonds van een uitgeverij. Dat neemt niet weg dat ook hier de verwachting is dat een classificatie mogelijk is. Die zal niet onmiddellijk volledige consensus opleveren, maar wel helpen een gedragsverandering te bewerkstelligen. Uitgevers zullen met redacties voor bepaalde series gaan werken, waardoor de in de tot die series behorende boeken gepubliceerde hoofdstukken goed classificeerbaar worden. Iets vergelijkbaars kan gelden voor monografieën, als die ook tot series behoren.

Indicatoren voor maatschappelijke kwaliteit

Binnen het ERiC-project werken KNAW, NWO, VSNU en de HBO-raad samen aan methoden voor de beoordeling van de maatschappelijke kwaliteit van wetenschap. Daar toe behoort ook de ontwikkeling van indicatoren. Daarnaast is er ook de landelijke commissie valorisatie (LCV). Het ligt voor de hand dat bij de verdere ontwikkeling van dit systeem wordt samengewerkt met beide projectgroepen. Dit is in overeenstemming met hetgeen is aangegeven in de *Strategische Agenda* van OCW (september 2011, pp. 52-53).

Het lijkt bovendien verstandig om op dit punt samen te werken met de KNAW-commissies die vergelijkbare exercities doen of hebben gedaan op het gebied van de technische en construerende wetenschappen en de sociale wetenschappen.

Afstemming methoden ontwerpende en construerende wetenschappen en sociale wetenschappen

Afgezien van het gezamenlijk belang van het ontwerpen van goede en geschikte indicatoren voor de maatschappelijke kwaliteit van wetenschappelijk onderzoek is er de vraag gesteld door het bestuur van de KNAW naar de overeenkomsten tussen de verschillende aanpakken. Deze vraag kwam ook aan de orde in het rectorenoverleg (zie noot 3). Ten aanzien van dit punt is er al een overleg geweest tussen de voorzitters van de diverse KNAW-projecten waarbij ook de voorzitter van het ERiC-project aanwezig was (voorjaar 2011). Het verdient aanbeveling dit traject opnieuw op te pakken en voort te zetten.

Verduidelijking extended peer review, rol van experts

In het voorgestelde systeem is er weinig aandacht besteed aan de functie en samenstelling van de *review*-commissie. Duidelijk is wel dat deze commissie een zeer centrale rol speelt omdat daar alle informatie die wordt verzameld via de indicatoren en criteria samenkomt en zorgt voor de onderbouwing van de uiteindelijke oordelen van de commissie. En duidelijk is ook dat vanwege het brede karakter van de beoordeling er in veel gevallen een bredere commissie gewenst is dan een traditionele *peer*-commissie. Er zijn verschillende manieren om dat te organiseren en een verdere verkenning van ervaringen op dit gebied is noodzakelijk.

NWO-pilot

Door het bureau van NWO wordt een evaluatie van het (reeds afgesloten) programma 'The Future of the Religious Past' voorbereid. Het betreft een traject waarin ons voorstel voor kwaliteitsindicatoren expliciet wordt meegenomen. Het wordt zowel inhoudelijk als op het niveau van de bestuurlijke aspecten beschouwd. Er zal overleg plaats vinden tussen het secretariaat van onze commissie en NWO betreffende de resultaten van de *pilots* bij het Meertens Instituut en ICOG en de ervaringen van het NWO-programma. Het Gebiedsbestuur Geesteswetenschappen is betrokken bij het hele proces van deze evaluatie. Een belangrijk onderwerp zal hier zijn de mogelijkheden om het systeem *ex ante* toe te passen bij de indiending van subsidievoorstellen voor onderzoeksprojecten.

Afstemming met CWTS-studie

CWTS voert in opdracht van het CvB van de Erasmus Universiteit Rotterdam (EUR) een studie uit naar de impact van het onderzoek zoals dat aan de EUR wordt bedreven, en geregistreerd in METIS. Het doel van de studie is tweërlei: (a) een inventarisatie van de output en impact van de verschillende onderdelen van de EUR zelf, en (b) een vergelijking van de output en impact van rechtsgeleerdheid en reschiedenis met die vakgebieden aan andere Nederlandse universiteiten.

Europese samenwerking

Ten slotte is het project op verschillende internationale (Europese) fora gepresenteerd, onder meer in het Europese RTD *evaluation network* en op verschillende conferenties gewijd aan indicatoren voor de geesteswetenschappen. Hoewel daaruit nog geen concrete stappen zijn voortgekomen voor verdere samenwerking is er wel regelmatig overleg met de Vlaamse collega's van het Vlaams Academisch Bibliografisch Bestand – VABB. Een interessante ontwikkeling daar is het kwaliteitslabel dat men bij uitgeverijen wil invoeren, het zogenoemde GPRC (Guaranteed Peer Reviewed Content). Het is een kwaliteitslabel waarmee de Vlaamse Uitgevers Vereniging (VUV) aangeeft dat de publicatie die dit label draagt, een *peer-review* procedure heeft doorlopen die beantwoordt aan de internationale wetenschappelijke standaarden. Dit is zeker iets waar naar moet worden gekeken bij de ontwikkeling van een classificatie van uitgeverijen.

Het tweede spoor: het bestuurlijke proces

Met de aflevering van dit eindrapport is het werk dat de commissie heeft gedaan, zoals gezegd, nog niet afgerond. Gebleken is dat er nog een flink aantal zaken ligt die verdere uitwerking behoeven alvorens van een volledig uitgebalanceerd systeem te kunnen spreken. En vervolgens moet worden nagedacht hoe dit te integreren in het nieuwe SEP. Bij dit laatste is het ook belangrijk te kijken naar de mogelijkheden om de benodigde data op te nemen in het vigerende landelijke systeem METIS (of een opvolger daarvan). Om te voorkomen dat er vertraging optreedt in de voortgang van dit veranderingsproces zal een beslissing moeten worden genomen over de regie van dat proces. Ervan uitgaande dat de huidige commissie met de aflevering van dit eindrapport zijn werkzaamheden beëindigt zal moeten worden bedacht hoe dit proces verder vorm kan worden gegeven. Daarbij kan worden geleerd van de ervaringen bij de 3 TUs, waar momenteel het systeem dat is voorgesteld door de KNAW-commissie voor construerende en ontwerpende wetenschappen, in de praktijk wordt uitgetest. Maar daarnaast zullen met diverse andere partijen afspraken moeten worden gemaakt. Te denken valt aan de decanen, de directeurs van de onderzoekscholen, NWO, de VSNU en het CWTS.

De commissie stelt zich voor dat een in te stellen stuurgroep daartoe het geëigende middel kan zijn. In het volgende hoofdstuk werken wij deze gedachte uit in het eigenlijke advies aan het bestuur van de KNAW.

4. ADVIES AAN HET BESTUUR VAN DE KNAW

De hoofdconclusie uit de pilotstudies en de veldraadpleging in 2011 is dat het door de KNAW-adviescommissie voorgestelde evaluatiesysteem brede ondersteuning ondervindt in het geesteswetenschappelijk veld, zowel bij onderzoekers als bestuurders. Er is consensus over het feit dat de commissie er in is geslaagd een systeem op te zetten dat enerzijds voldoende flexibiliteit kent om de variatie in vakgebieden te representeren en anderzijds de mogelijkheid biedt om eenheden onderling te vergelijken. Er bestaat waardering voor het feit dat er een evenwicht is gevonden voor de beoordeling van de wetenschappelijke kwaliteit en de maatschappelijke relevantie. En het feit dat de *review*-commissie (in overleg met de opdrachtgever) verantwoordelijk is voor de weging van de verschillende indicatoren wordt algemeen gezien als een groot pluspunt. Kritische kanttekeningen werden ook gemaakt, ze zijn hierboven uitgebreid aan de orde gekomen (hoofdstukken 2 en 3).

Nu komt het er op aan het proces verder te trekken. Het is gelukkig zo dat het overleg van de decanen letteren en geschiedenis (DLG) dermate enthousiast is over het voorgestelde systeem dat ze hebben aangegeven het voortouw te willen nemen inzake de verdere implementatie. Onze commissie heeft daarom de volgende twee aanbevelingen aan het bestuur van de KNAW:

1. Vraag het DLG om samen met de twee andere disciplineoverleggen in de geesteswetenschappen (theologie en wijsbegeerte) en de onderzoekscholen te bevorderen dat dit systeem wordt geïmplementeerd in de reguliere evaluatiepraktijk (in de eerste plaats het SEP, maar ook andere waar mogelijk). Prioriteit in dat proces heeft de totstandkoming van een goede tijdschriften-*ranking*.

2. Stel een stuurgroep in die de verdere uitwerking van het systeem bevordert. Deze stuurgroep zou kunnen bestaan uit enkele leden van de huidige commissie (Keimpe Algra en Hans Bennis), de directeur van het CWTS (Paul Wouters), een vertegenwoordiger van de decanen geesteswetenschappen, en de voorzitter van het ERiC-project. De secretaris van de huidige adviescommissie zou ook de secretaris van de stuurgroep moeten zijn. De taken van deze stuurgroep zijn:
- Het gebruik van dit systeem binnen de geesteswetenschappen in den brede te bevorderen: bij faculteiten, onderzoekscholen, instituten, NWO-aanvragen.
 - De discussie over deze brede vorm van evaluatie in de adviesraden van de KNAW en breder in het veld, te bevorderen. Dit mede in het licht van het feit dat er nu drie adviesrapporten zijn (worden) uitgebracht over deze problematiek. Toewerken naar een conferentie/workshop zo mogelijk met NWO en VSNU zou daarbij voor de hand liggen.

De commissie heeft goede hoop dat de stuurgroep in het vervolgtraject kan profiteren van een aantal verwante ontwikkelingen:

- De ervaringen van de 3 TUs, die een protocol hebben ontworpen waarin plaats is ingeruimd voor de bevindingen van het rapport over de technische en construerende wetenschappen.
- Het ERiC-project waarin KNAW, VSNU, NWO, HBO-raad en Rathenau Instituut samenwerken aan de ontwikkeling van indicatoren voor de maatschappelijke relevantie. Met een vorig jaar uitgebrachte handreiking wordt op diverse plaatsen geëxperimenteerd.
- De Landelijke Commissie Valorisatie waarin de KNAW eveneens is vertegenwoordigd, via de president, waarin wordt gewerkt aan de ontwikkeling van indicatoren voor de valorisatie van wetenschappelijk onderzoek in den brede.
- De bevindingen van twee andere adviesraden van de KNAW: het rapport *Kwaliteitsbeoordeling in de Ontwerpende en Construerende Disciplines* en het werk van de commissie Bensing voor de sociale wetenschappen.
- Het CWTS voert momenteel een studie uit bij twee faculteiten van de Erasmus universiteit (rechten en geschiedenis) waar de mogelijkheden voor bibliometrische indicatoren worden onderzocht.
- NWO voert dit jaar een pilotstudie uit rond het (afgesloten) programma 'The Future of the Religious Past'. Hierbij wordt het door ons voorgestelde systeem als een richtsnoer gebruikt.

BIJLAGE 1

DE PILOTSTUDIE

Inleiding

De Commissie Kwaliteitsindicatoren Geesteswetenschappen (KIG) heeft in juni 2011 een interim-adviesrapport uitgebracht waarin een systeem wordt voorgesteld om de kwaliteit van het onderzoek in de geesteswetenschappen in zijn volle breedte te kunnen beoordelen. De commissie komt tot de conclusie dat de geesteswetenschappen niet fundamenteel verschillen van andere disciplines maar wel andere accenten leggen voor wat betreft producten en publicatiekanalen. De commissie KIG stelt een systeem voor van kwaliteitsindicatoren waarbij zowel de wetenschappelijke als de maatschappelijke kwaliteit kan worden beoordeeld, met gebruikmaking van indicatoren verdeeld over drie categorieën, output, gebruik en erkenning.

Als vervolg op het rapport heeft de commissie besloten een *pilot* te doen ten behoeve van de verdere ontwikkeling van indicatoren. De pilot kent twee hoofddoelstellingen:

1. Vaststellen of het voorgestelde scala aan indicatoren herkend wordt door de onderzoekers en een voldoende representatief beeld geeft van de inspanningen, output en resultaten van geesteswetenschappelijk onderzoek.
2. Vaststellen in hoeverre voor de voorgestelde indicatoren data beschikbaar zijn, en zo niet of deze met een redelijke inspanning zijn te verzamelen.

De *pilot* bestond uit de volgende onderdelen:

- Casestudies bij het Instituut voor Cultuurwetenschappelijk Onderzoek Groningen (ICOG) van de Rijksuniversiteit Groningen en het Meertens Instituut van de KNAW.

- Workshop voor het hele veld van de geesteswetenschappen, gehouden op 25 november 2011 bij het NIAS.
- Rondetafelgesprek met beleidsmedewerkers en enkele onderzoekers die verantwoordelijk zijn voor de evaluatie van hun instituut (26 september 2011).

De twee hoofddoelstellingen zijn vertaald in de volgende vragen:

- a. het voorgestelde systeem representatief voor het veld en herkenbaar voor de onderzoekers?
- b. zijn de voorgestelde indicatoren werkbaar in de praktijk, en zijn er robuuste data beschikbaar of te verzamelen met een redelijke inspanning?
- c. hoe verhouden de indicatoren zich tot de categorieën binnen het bestaande registratiesysteem METIS?

Om deze vragen te beantwoorden zijn in de twee genoemde cases relevante documenten bestudeerd (jaarverslagen, evaluaties) en er zijn enkele kwantitatieve analyses uitgevoerd gericht op het in beeld brengen van de output van groepen. Ook zijn er gesprekken gevoerd met de directies van beide instituten. In verschillende focusgroepen is met onderzoekers en informatiespecialisten gediscussieerd over de bruikbaarheid van het systeem voor de eigen onderzoeks- en publicatiepraktijk, over welke indicatoren zij van belang vonden en over de mogelijke voorstellen die zij hadden om deze verder uit te werken. De focusgroepen waren zo veel mogelijk multidisciplinair samengesteld. Onderzoekers kwamen vanuit de linguïstiek, etnografie, historische economie, letteren (diverse talen), kunstgeschiedenis, cultuurwetenschappen, journalistiek en media. Bij de samenstelling van de focusgroepen is ook gelet op verscheidenheid in rangen. Zo namen hoogleraren, UHD's, UD's en promovendi deel aan de gesprekken. Aan een van de focusgroepen namen ook geesteswetenschappelijke onderzoekers van elders deel.

In de groepen zijn de hierboven genoemde vragen besproken. Daarnaast is met het oog op het aspect van de dataverzameling en registratie nog een aparte discussiebijeenkomst georganiseerd voor beleidsmedewerkers van faculteiten geesteswetenschappen en verantwoordelijke bestuurders. De voorlopige uitkomsten van de gesprekken en inventarisaties zijn gepresenteerd en besproken op de workshop van 25 november jl. bij het NIAS, waaraan onderzoekers, beleidsmedewerkers en bestuurders uit diverse gremia van de geesteswetenschappen deelnamen. Parallel aan onze pilotstudie heeft het CWTS in opdracht van de EUR onderzoek gedaan naar verschillen in publicatieculturen en de betekenis daarvan voor de mogelijkheden en de beperkingen van bibliometrische analyses voor de vakgebieden geschiedenis en recht. Met de onderzoekers van het CWTS is regelmatig overlegd over wederzijdse bevindingen. In het onderstaande rapporteren wij over de pilotstudies bij de twee instituten, de bijeenkomst met de beleidsmedewerkers en bestuurders en de NIAS-workshop. Daarna volgen de conclusies uit het geheel.

De verslagen van de case studies volgen de drie bovengenoemde vragen.

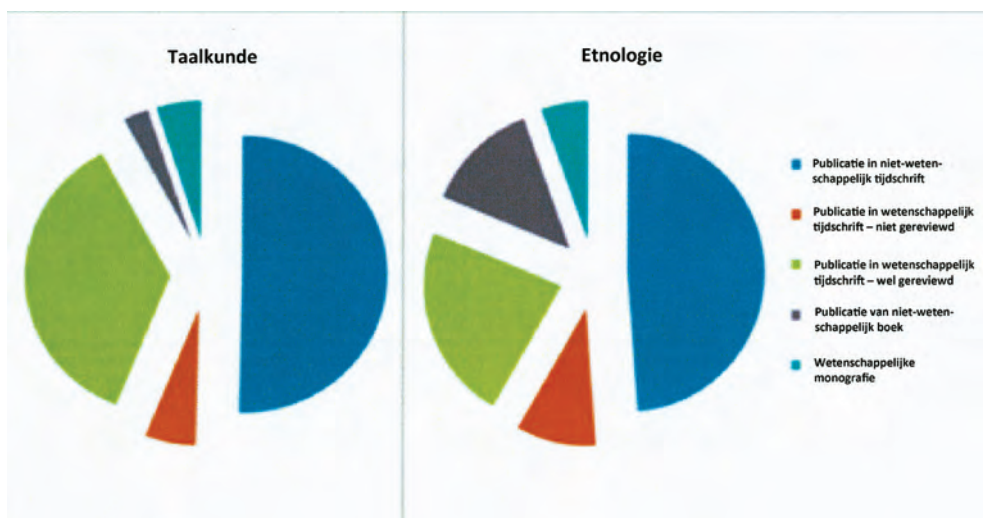
Casestudie bij het Meertens Instituut

a. Representativiteit van het voorgestelde systeem

Er zijn bij het Meertens twee focusgroepen gehouden. Deelnemers aan beide groepen vinden het voorgestelde systeem over het algemeen goed herkenbaar. Wat daarbij helpt is de aandacht voor specifieke output van de geesteswetenschappen (boeken!) en ook de aandacht voor de maatschappelijke output van onderzoek. Wel merken de deelnemers op dat het voorstel met name op het punt van die maatschappelijke kwaliteit niet voldoende expliciet is. De relatie tussen maatschappelijke en wetenschappelijke kwaliteit is in beide groepen een belangrijk onderwerp. Er wordt opgemerkt dat het lijkt of het systeem een impliciete hiërarchie aanbrengt tussen wetenschappelijke en maatschappelijke kwaliteit. Dat zou een probleem kunnen vormen voor de weging tussen beide aspecten, temeer daar de robuustheid van de data nogal verschilt. Bijvoorbeeld externe en sociale aspecten van onderzoek, zoals congressen, tentoonstellingen en ook de doorwerking van wetenschappelijk onderzoek in sociale en andere media zijn belangrijke indicaties voor de maatschappelijke relevantie maar data hiervoor zijn over het algemeen minder robuust dan de data voor wetenschappelijke kwaliteit (en worden soms ook minder goed bijgehouden). In deze discussie werd ook de relatie van een evaluatie tot de missie van het onderzoek een belangrijke factor genoemd. Het ene onderzoek is nu eenmaal meer gericht op maatschappelijke vragen dan het andere. Om recht te doen aan die verschillen en een goede afweging mogelijk te maken is het noodzakelijk daarmee rekening te houden bij de samenstelling van beoordelingscommissies. Het voorgestelde systeem laat echter nog niet zien hoe een commissie moet worden samengesteld.

De effecten van (sub)disciplinaire verschillen kwamen nadrukkelijk aan de orde. Taalkundigen vinden bijvoorbeeld dat zij een voornamelijk wetenschappelijke missie hebben en dat zij weinig hebben aan indicatoren over het maatschappelijk effect van hun werk. Zij zijn meer dan de etnologen gericht op internationale tijdschriftpublicaties, en de vraag is dan hoe dat in de afweging van de *review*-commissie uitwerkt. We hebben de output van beide groepen onderzocht en de indruk van de taalkundigen wordt deels bevestigd door de kwantitatieve analyse van recente publicatiegegevens. In verhouding tot het totaal van enkele van de belangrijkste publicatiecategorieën blijken linguïsten meer in de internationale tijdschriften (*peer reviewed*) te schrijven dan etnologen. De etnologen publiceren in verhouding wat meer monografieën. Echter, voor beide groepen geldt ze voor wat betreft het aandeel publicaties in niet-wetenschappelijke media ongeveer even hoog scoren (zie figuur 1). Op grond van deze gegevens is de conclusie dat de verschillen minder groot zijn dan sommigen

denken, en dat er geen reden is te vrezen voor moeilijkheden bij de afweging in de *review*-commissie.



Figuur 1 Relatieve aandelen van aantallen voornaamste publicatievormen voor twee specialismen van het Meertens Instituut, data gebaseerd op METIS-gegevens 2006-2010

b. Werkbaarheid en robuustheid van de indicatoren

Deelnemers aan de focusgroepen hebben twijfels bij de vraag of de indicatoren voldoende robuust zijn. Zij menen dat men er niet zomaar mee aan de slag kan: de indicatoren in het schema zijn veeleer categorieën van indicatoren. Een nadere uitwerking is noodzakelijk en pas dan kan men zien in hoeverre bijvoorbeeld de diversiteit in publicatievormen of nieuwere vormen van output, op een goede manier ‘gevangen’ wordt door dit voorstel. Maar deelnemers waarschuwen ook voor een te grote proliferatie aan indicatoren die tot een minder werkbaar systeem zal leiden. Het al eerder vermelde verschil in robuustheid tussen indicatoren voor de wetenschappelijke en maatschappelijke kwaliteit kan volgens de deelnemers leiden tot onduidelijkheden in de *review*-commissie. Immers, zo lang de *reviewers* geen duidelijke aanwijzingen hebben hoe beide aspecten af te wegen zullen de indicaties voor wetenschappelijke kwaliteit de boventoon voeren, zo is de verwachting.

Terwijl men beducht is voor een te grote hoeveelheid indicatoren, worden er toch ook voorstellen gedaan voor toevoeging van indicatoren. In het bijzonder zou men meer zichtbaarheid willen voor het internationale aspect, bv. door copublicaties, samenwerking in projecten e.d. te overwegen als indicatoren. De commissie acht deze suggesties waardevol, maar is van mening dat ze, indien er consensus over bestaat in het veld of subveld, kunnen worden ingebracht in de categorie ‘overig’.

De vraag of het mogelijk is voor het instituut te komen tot een breed geaccepteerde *ranking* van tijdschriften met bv. A-, B- en C-categorieën levert een discussie op over twee aspecten, ten eerste de grote variatie aan tijdschriften en uitgeverijen in de output van het Meertens, en ten tweede het taal-aspect. De variatie doet zich voor tussen de beide vakgebieden (er blijkt weinig overeenstemming te bestaan tussen beide gebieden voor wat betreft de tijdschriften), maar ook binnen de disciplinaire groepen is een aanzienlijke variatie. Zo blijken voor taalkunde 47 artikelen in *peer-reviewed* tijdschriften verspreid te zijn over 14 tijdschrifttitels en bij etnologie 28 publicaties over 23 verschillende titels. Voor de 47 publicaties van taalkunde in de vorm van boeken en hoofdstukken in bundels vonden zij 19 verschillende uitgeverijen. De 33 publicaties van etnologie in boekvorm verschenen bij 22 verschillende uitgeverijen. De conclusie van de aanwezige onderzoekers is dat het niet eenvoudig zal zijn om tot overeenstemming te komen over één lijst van tijdschriften of uitgeverijen. Daar komt nog bij dat er onzekerheid bestaat over de verhoudingen tussen de internationale en nationale domeinen. Dat geldt in de eerste plaats de betekenis van het publiceren in het Nederlands. Men verwijst naar het KNAW-rapport *Nederlands, tenzij* (2003) en vraagt ‘geldt dat nog’? In dat rapport wordt een pleidooi gehouden voor het behoud van Nederlands als wetenschapstaal voor bepaalde domeinen. Voor wat betreft het internationale domein geldt dat weliswaar het Engels nu dominant is, maar in de geesteswetenschappen zijn Duitse, Franse en in toenemende mate ook Aziatische media van belang. De commissie zit niet op het spoor van *Nederlands, tenzij*, maar zij ziet het ook niet als haar taak prescriptief Engels als publicatietaal op te leggen. De commissie neemt in deze kwestie geen normatief standpunt in, maar baseert zich op de feitelijke praktijk.

Deelnemers wijzen er voor wat betreft een eventuele *ranking* van uitgeverijen op dat bij het bepalen van de kwaliteit van uitgeverijen het al of niet hanteren van peer review een onbetrouwbaar criterium kan zijn. *Peer-review*-praktijken van uitgeverijen kunnen onderling sterk verschillen, en zelfs binnen uitgeverijen kan dat het geval zijn. Dat varieert van de eigen redactie (vaak bij Frans- en Duitstalige uitgeverijen) tot een werkelijk internationaal *peer review* (Angelsaksisch).

Ten slotte wordt opgemerkt dat de uitwerking van het gebruik van bepaalde indicatoren voor uitgeverijen en tijdschriften gevolgen kan hebben, sommige gunstig, andere minder gunstig. Bestaande *review*-procedures zullen wellicht worden verbeterd, bijvoorbeeld door de procedure meer onafhankelijk of internationaler te maken, wat een gunstig effect zou kunnen hebben op de kwaliteit van de review en uiteindelijk het tijdschrift. Anderzijds kan het ook leiden tot het verdwijnen van tijdschriften die niet aan de nieuwe normen willen of kunnen voldoen (zoals dat bijvoorbeeld in Vlaanderen het geval lijkt te zijn na invoering van indicatoren voor de humane wetenschappen).

- c. Beschikbaarheid van de informatie voor het nieuwe systeem
Voor het vastleggen van informatie over publicaties gebruikt het Meertens instituut METIS, maar dat is uitgebreid met een eigen MONITOR-systeem. De meeste vormen van output die in het KIG-systeem worden voorgesteld komen al voor in dat systeem. Maar informatie over de criteria 'gebruik' en 'erkenning' ontbreekt grotendeels. In sommige gevallen worden categorieën niet gebruikt (bv. cds, dvds), in andere is er geen plek voor (tentoonstellingen, tekstedities). En zo lang informatie niet kan of hoeft te worden ingevuld ontbreekt bij onderzoekers de prikkel om dit te melden. Er is dus sprake van een probleem dat zowel technische kanten heeft als bestuurlijke. Om een evenwichtige beoordeling te verkrijgen is het noodzakelijk om een balans te vinden voor de te verzamelen informatie voor de verschillende in het KIG-systeem voorgestelde indicatoren.

Instituut voor Cultuurwetenschappelijk Onderzoek Groningen (ICOG)

- a. Representativiteit van het voorgestelde systeem
Het merendeel van de ICOG-medewerkers vindt dat het voorgestelde systeem goed aansluit bij wat zij als een centraal kenmerk van het veld zien: een arena van inzichten met zowel wetenschappelijke als maatschappelijke betekenis. Hoewel de reacties ook bij het ICOG over het algemeen positief zijn, lopen ze meer uiteen dan bij het Meertens Instituut. Er is ook hier waardering voor de aandacht voor de monografie, niet alleen omdat het de belangrijkste publicatievorm is maar vaak ook een voorwaarde voor het binnenhalen van subsidie.

De deelnemers bij het ICOG zijn afkomstig uit meer verschillende disciplines dan het bij het Meertens. Verschillen tussen die disciplines lijken echter ook hier meer in relatie te staan tot particuliere wetenschapsopvattingen dan tot verschillen in disciplinaire praktijken of missies.

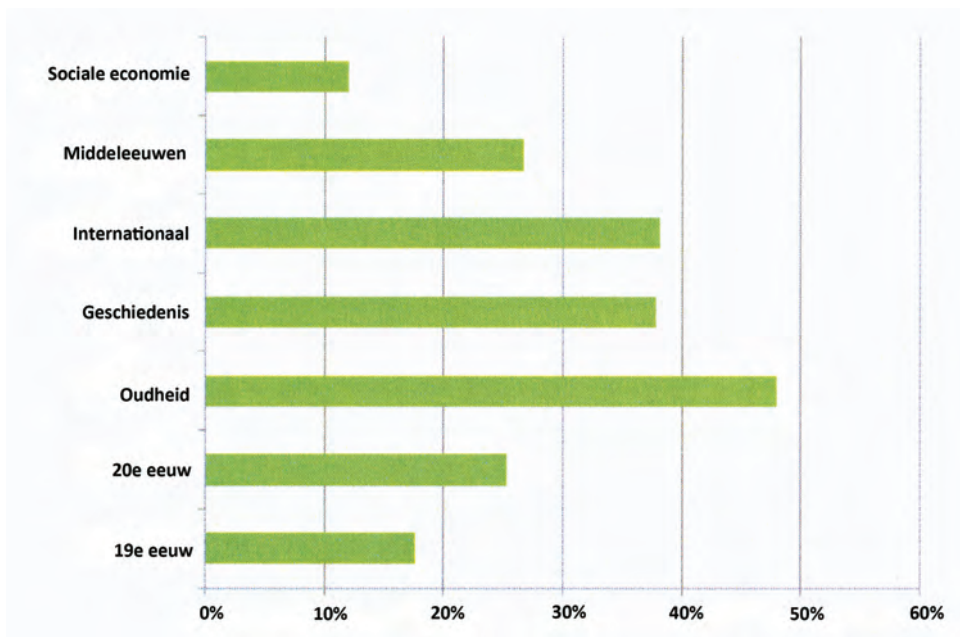
Een minderheid van de deelnemers ziet het produceren van maatschappelijke output als hooguit een afgeleide activiteit (vooral de letterkundigen, maar ook weer niet alle letterkundigen). Iemand noemt het een bureaucratisch middel om grip te krijgen op de onderzoekspraktijk en een systeem met allerlei toegevoegde categorieën die de gelegenheid bieden te verbloemen van wat gezegd zou moeten worden over wetenschappelijke kwaliteit. Maar het merendeel van de deelnemers is het er over eens dat wetenschappelijke en maatschappelijke oriëntatie beide belangrijk zijn voor geesteswetenschappelijk onderzoek. Dat hoeft echter niet te betekenen dat het voor alle individuele onderzoekers in gelijke mate verplichtend is om maatschappelijk relevant onderzoek te produceren. Dat heeft ook als gevolg dat bij beoordelingen de weging van wetenschappelijke en maatschappelijke output globaal moet zijn, bijvoorbeeld op het niveau van de onderzoeksgroep. De meerderheid van de deelnemers vindt wel dat in de beoordeling wetenschappelijke kwaliteit zwaarder moet wegen dan de maatschappelijke kwaliteit. Maar het is onvoldoende helder hoe de afweging tussen beide aspecten

moet worden gemaakt, hoe dwingend bijvoorbeeld een *mission statement* is bij het vellen van een oordeel en vooral hoe dat uitwerkt bij vergelijkende oordelen.

Er wordt ook opgemerkt dat wetenschappelijke en maatschappelijke verrichtingen elkaar wederzijds kunnen versterken en inspireren. Het is heel goed mogelijk dat wetenschappelijke output, zoals een boek, een cross-over kan zijn met maatschappelijke output, dat wil zeggen dat die wetenschappelijke output tevens tot maatschappelijke relevante resultaten kan leiden. Dat geldt ook andersom, wanneer bijvoorbeeld een catalogus voor een publiekstentoonstelling een wetenschappelijke effect bereikt.

b. Werkbaarheid en robuustheid van de indicatoren

Ook hier veel discussie over de tijdschriften-*ranking* en het eventuele onderscheid van tijdschriften en uitgeverijen in A-, B- en C-categorieën. Er wordt een principieel bezwaar gemaakt: dergelijke onderscheidingen zijn gebaseerd op wat in de economie 'inputmodellen' heten en ze leiden uiteindelijk tot ongewenst gedrag. Het beste zou zijn een beoordeling louter te baseren op eindresultaat, zoals de impact die met Web of Science gemeten kan worden vanuit de gedachte dat vooraanstaande artikelen en boeken altijd zichtbaar worden in dergelijke databases. Maar de meesten zijn het daarmee niet eens, vooral omdat er grote verschillen zijn tussen gebieden in de geesteswetenschappen. De bijgaande grafiek laat de percentages zien van tijdschriftpublicaties van verschillende ICOG vakgebieden die in Web of Science als bron gebruikt worden.



Figuur 2: Percentage wetenschappelijke artikelen dat in WoS-tijdschriften verschijnt, data gebaseerd op METIS-gegevens 2004-2009

Uit de grafiek blijkt dat sommige gebieden veel meer gericht zijn op de internationale gemeenschap (bv. oudheid) en andere op het lokale discours (bv. sociaal economische studies). De vraag is in hoeverre dit het gevolg is van bewust publicatiebeleid of dat er andere oorzaken zijn die deze verschillen kunnen verklaren.

Uit de kwantitatieve analyse van publicatiegegevens blijkt voorts dat disciplines niet zonder meer de verbindende factor te zijn in de keuze die onderzoekers maken bij het publiceren in wetenschappelijke tijdschriften. Ten tijde van het onderzoek telde ICOG negentien onderzoeksgroepen met uiteenlopende disciplines. Weliswaar deelden de onderzoekers uit deze groepen sommige tijdschriften, zodat deze titels gezien zouden kunnen worden als gemeenschappelijke platforms voor academische uitwisseling. Maar het gaat daarbij slechts om drie tijdschriften waarin ICOG medewerkers tien keer of meer publiceerden, en daarnaast zijn er vijftien verschillende tijdschriften waarin zij vijf keer of meer publiceerden. De vijfhonderd wetenschappelijke artikelen van ICOG verschenen in 284 verschillende tijdschrifttitels, waarbij 75% van deze tijdschriften slechts één maal voorkwamen in de lijst, en 86% een of twee maal.

Het is uiteraard lastig om met zo'n divers bestand aan tijdschriften tot consensus over een *ranking* te komen, tenzij misschien in heel kleine gebieden. In de discussie over classificatie van tijdschriften merken enkele onderzoekers op dat zij hun publicaties spreiden over verschillende wetenschappelijke en maatschappelijke publiek. Het zou daarom wenselijk zijn bij eventuele classificatie een onderscheid te maken naar specialistische en generalistische tijdschriften (zoals dat ook gebeurt in de ERIH-lijst van de ESF). Ook wordt gewaarschuwd voor het gevaar dat er een systeem wordt ontworpen waardoor waardevolle Nederlandse tijdschriften verdwijnen, iets wat het gevolg kan zijn van een te sterke nadruk op internationalisering. Er wordt verwezen naar effecten van die druk in Vlaanderen en ook in Noorwegen (voor Noorstalige tijdschriften).

Er wordt bij de discussie over criteria en indicatoren voor publicaties gewezen op de effecten op het onderzoeksbeleid van groepen. Recentelijk is bijvoorbeeld de publicatiestrategie van individuele medewerkers onderdeel gemaakt van het persoonlijk begeleidingsgesprek. In dat gesprek wordt de keuze voor tijdschrifttitels besproken en de inspanning om (alleen of vooral) te publiceren in congresbundels ontmoedigd – tenzij het congres ook werkelijk statuut heeft in het veld. Ook komt de grote variatie in kwaliteit van tijdschriften ter sprake. Er is dus sprake van een strategisch effect van deze discussie.

Er is consensus over het feit dat de kwaliteit van tijdschriften en uitgeverijen bepaald wordt door de aanwezigheid van een *peer-review*-systeem. Maar men erkent ook dat er verschillende vormen van *peer review* zijn die niet allemaal even betrouwbaar zijn.

In aanvulling op de KIG-lijst van indicatoren wordt nog een aantal andere mogelijke indicatoren genoemd. Bij het criterium wetenschappelijke erkenning zou bijvoorbeeld ook gerekend kunnen worden *fellowships*, *keynotes* en prijzen.

Die blijken bijvoorbeeld in de VS aan belang te winnen als categorie. Daarnaast zouden ook *reviews* als indicator kunnen worden opgenomen. Het is belangrijk voor het vakgebied dat een dergelijke categorie erkenning krijgt. Ook het belang bundels mag niet worden onderschat omdat hier sprake is van een debat onder vakgenoten. Men realiseert zich wel dat de huidige registratie van *reviews* in METIS anders gedaan moet worden. Er komen grote aantallen reviews voor in de publicatielijsten van het ICOG, en deze zouden op zijn minst moeten worden gesplitst naar maatschappelijke publicaties en reviews in wetenschappelijke tijdschriften.

Er is ook discussie over het nut van citaties voor de geesteswetenschappen. Sommigen vinden dat het een rol zou moeten spelen om blijken van erkenning te kunnen meten, maar dan niet alleen via de WoS. Er is bovendien twijfel of medewerkers dat allemaal zelf kunnen en willen bijhouden (hoewel bv. NWO dat wel verlangt). Maar de verwachting is ook dat bestaande redelijk betrouwbare databases zoals *Web of Science* of Google Scholar een te geringe afspiegeling geven van de totale respons op geesteswetenschappelijk werk. Hoewel het een illusie is dat ooit alles bibliometrisch zichtbaar gemaakt kan worden is het van belang om zoveel mogelijk zichtbaar te worden voor de uiteenlopende databases: wat daarin niet terecht komt wordt ook niet van echt belang gevonden. Om die redenen wordt gepleit voor *open-access*-publicaties.

c. Beschikbaarheid van informatie voor het nieuwe systeem

Er blijkt ook hier sprake te zijn van een zekere *mismatch* tussen het METIS-systeem dat het ICOG hanteert en het KIG-systeem. Uit de inventarisatie bij ICOG blijkt dat de gehanteerde categorieën voor output in aantal beperkt zijn, ook in het licht van wat in METIS en volgens de VSNU-richtlijn mogelijk is. Zo maakt het ICOG geen onderscheid op grond van het al dan niet *gereviewed* zijn van publicaties (althans in de periode waarover dit onderzoek zich uitstrekt; inmiddels gebeurt dat wel). Ook valt op dat de categorieën nogal breed zijn, in het bijzonder die van de professionele publicatie: deze categorie kan tijdschriften van zeer verschillende aard en zwaarte omvatten, en dat geldt *mutatis mutandis* ook voor bundels en boeken.

De verantwoordelijken voor de invoer van informatie geven aan nogal eens problemen te hebben met het maken van de juiste keuzes in METIS. Het is niet altijd duidelijk of wat wordt aangeleverd onder categorie 'wetenschappelijk' valt of onder 'professioneel' of 'overig'. Enerzijds wordt dit in de hand gewerkt doordat onderzoekers onvoldoende op de hoogte lijken te zijn van de gewenste relevante informatie, en wellicht ook minder gemotiveerd lijken te zijn om informatie aan te leveren. Anderzijds functioneert het METIS systeem door de datavalidatie soms als een procrustesbed voor beschikbare gegevens.

Bij het aanleveren van de informatie voor METIS zouden de medewerkers een nadrukkelijker rol kunnen spelen. Er is op dit moment een gevaar van onderrapportage omdat medewerkers de informatie niet goed genoeg aanleveren.

Zelf invoeren is wel te doen, maar nu nog lastig. Medewerkers zouden bewust gemaakt moeten worden van het belang. Het voordeel van zelf doen is dat je de informatie kunt doorvoeren naar je eigen internetpagina (MEPA) en dat deze beschikbaar is voor collega's en anderen. Voor sommige disciplines wordt dat van groot belang geacht, bijvoorbeeld bij taalkundige onderzoekers omdat onderzoekers in India en China lang niet altijd toegang hebben tot een bibliotheek. In dat opzicht is het publiceren in open-access-tijdschriften ook van belang.

Workshop KIG, NIAS 25 november 2011

Voorafgaand aan de workshop hebben we aan vier vertegenwoordigers van de belangrijkste betrokken gremia gevraagd om een reactie op het door ons voorgestelde systeem. Die vatten we eerst kort samen. Daarna volgt nog een korte impressie van de discussie met de zaal.

Annemarie Korte (namens gebiedsbestuur NWO geesteswetenschappen)

Instemming met de voorgestelde aanpak die helder, transparant en eenvoudig wordt genoemd. NWO wil graag meewerken aan het project en stelt voor het programma 'Future of the Religious Past' langs de lijnen van het KIG-voorstel te evalueren. Daarbij zou ook moeten worden getest in hoeverre dit systeem nuttig is voor *ex-ante*-beoordeling van voorstellen. Gevraagd wordt waarom niet is gekozen voor een opsplitsing in een deel centrale indicatoren die voor alle vakgebieden gelden en een deel specifiek die per vakgebied kunnen gelden. De commissie antwoordt dat men er juist voor heeft gekozen dat niet te doen omdat wat voor het ene gebied geldt niet voor het andere hoeft te gelden.

Gerry Wakker (namens decanenoverleg letteren en geschiedenis)

De decanen zijn zeer positief over wijze waarop de diversiteit van de geesteswetenschappen wordt aangepakt. Het plan lijkt bovendien zeer uitvoerbaar. Wel moet de maatschappelijke kwaliteit nader worden uitgewerkt, ook om te kunnen gebruiken in de beleidsdiscussies en voortgangsgesprekken. Er moet goed worden nagedacht over hoe wordt vastgesteld welke indicatoren geschikt zijn bij de beoordeling van faculteiten. De decanen willen hier het voortouw nemen, maar ook de onderzoekscholen moeten hierbij betrokken worden. Datzelfde geldt *mutatis mutandis* voor het vaststellen van een onderscheid ABC-tijdschriften en uitgeverijen. Het systeem moet ook bijdragen aan mentaliteitsverandering, DLG wil dat decanen dit systeem actief uitdragen bij faculteiten en CvB's.

Marcus Düwell (LOGOS)

De onderzoekscholen zijn blij met het conceptuele onderscheid tussen indicatoren en criteria, dat maakt flexibiliteit mogelijk. Om te komen tot ERIH-achtige lijsten (op zich verstandig!) is het wel nodig een structuur voor feedback te organiseren, waarbij onderzoekscholen worden betrokken. Het is belangrijk daarvoor een goed aggregatieniveau te kiezen, bij voorkeur niet het vakgebied maar het domein. Dat zou landelijk moeten worden vastgesteld, waarbij het de vraag is wie dat op een gezaghebbende wijze kan doen.

Jozien Bensing (KNAW commissie Indicatoren Sociale Wetenschappen)

De decentrale vaststelling van indicatoren is een goed punt omdat niemand meer kan

zeggen dat er geen ruimte is voor eigen benadering. Maar het probleem is wel hoe dat te operationaliseren. Je kan wel epistemologische verschillen tussen wetenschapsgebieden aangeven, maar er zijn ook nog andere verschillen die tijdsgebonden, generatiegebonden zijn. Jongere onderzoekers doen bv. al veel digitaal, maken gebruik van Facebook, Twitter. Daarbij komen oude vragen in een nieuw daglicht te staan: wat is de rol van *peer review* in digitale tijdperk, op welk moment is een publicatie ook een publicatie (is een blog een publicatie?). Digitalisering leidt tot meer dynamische productie, interactiever (dus ook minder goed toe te kennen aan een individu), vluchtiger ook. De meting is nu gericht op 'stollingsmomenten' maar dat houdt geen rekening met de nieuwe dynamiek in de wetenschappelijke gemeenschap.

Discussie met de zaal (een korte impressie)

Over criteria

Aandacht wordt gevraagd voor hoe criteria vergelijkend kunnen worden ingezet, omdat bij NWO in een aantal gevallen over alle disciplines wordt beoordeeld: bij VICI bv. moet een afweging gemaakt worden tussen aanvragen van alfa, bèta, gamma en medisch. Het probleem blijft dat vergelijking tussen vakgebieden lastig is vanwege verschillende publicatie- en communicatietradities. Voor anderen is het vanwege die verschillen alleen maar zinvol om criteria per discipline op te stellen. Het CWTS doet momenteel onderzoek naar verschillen in publicatiecultuur, o.a. bij geschiedenis en rechten. Speciale aandacht is daarbij nodig voor interdisciplinair werk dat in het nadeel is bij het opstellen van interne (facultaire) ranglijsten. Daarom zou men bij het opstellen van classificaties een aggregatieniveau moeten kiezen dat de faculteiten overstijgt. Voorgesteld wordt om lijsten die nu al her en der circuleren te bekijken en te bezien wat daarin de overeenkomsten zijn en wat voor effecten die lijsten hebben. Dit voorstel ondervindt veel bijval. In deze discussie over classificatie van tijdschriften wordt ook nog gewezen op het belang van tijdschriften die van belang zijn voor de (nieuwe) ontwikkelingen in een vakgebied (dus niet alleen letten op de status quo/gevestigde orde).

Over indicatoren

Er worden diverse suggesties gedaan voor toevoeging van indicatoren. Voorbeelden zijn databases, *peer review* van boeken, (het schrijven van) recensies. Er wordt op gewezen dat er weinig aandacht is voor nieuwe ontwikkelingen, met name rond de digitale mogelijkheden van *open access*. Hierdoor hebben media een veel groter bereik, het lezen via *open access* is inmiddels al een veelvoud van het uitlenen van boeken. Publiceren via de nieuwe media zou in de ogen van sommigen beloond moeten worden. Het antwoord van de commissie is dat het systeem over voldoende flexibiliteit beschikt om nieuwe ontwikkelingen toe te laten. In de categorie 'overig' is ruimte voor nieuwe indicatoren. Het moet natuurlijk wel mogelijk zijn om daarvoor data te verzamelen.

Meerdere deelnemers maken zich zorgen over de gevolgen die teveel nadruk op indicatoren kan hebben voor interdisciplinair en nieuw onderzoek. Indicatoren kijken

terug en zijn vaak niet gericht op het kijken over de grenzen van vakgebieden. Er wordt voor gepleit de onderzoeksscholen te betrekken bij deze problematiek omdat die vaak interdisciplinair georiënteerd zijn. Zij hebben ook minder last dan faculteiten van andere besommingen (financiële en personele zaken) die belemmerend kunnen werken voor interdisciplinair onderzoek.

De globale lijn van het systeem wordt ook geprezen. Voorkomen moet worden dat men te ver gaat en aan micromanagement gaat doen. Als je te ver doorschiet is het risico groot dat je dat je de onderzoekers te dicht op de huid zit en dat de effecten contraproductief zijn. Het systeem zou niet moeten ontaarden in een afrekenmethodiek, maar gericht op verbetering van het onderzoek en het onderzoeksbeleid (bijvoorbeeld: het carrièreperspectief).

Over METIS

Er zijn twee zaken die naar voren komen. In de eerste plaats moet METIS beter aansluiten bij de huidige onderzoekspraktijk en de informatie-eisen die in het KIG-systeem worden gesteld. Er zijn in METIS weliswaar mogelijkheden om categorieën toe te voegen, maar er zou meer systematisch gekeken moeten worden naar welke informatie nodig is om daarop nieuwe (standaard)categorieën te baseren. In de tweede plaats is het nu zo dat universiteiten elk afzonderlijk beslissen over de invulling, dat helpt niet bij de ontwikkeling van een landelijk systeem. Men zou afspraken moeten maken met de colleges van bestuur om het systeem meer te uniformeren.

Hoe verder

De belangrijkste vraag aan het einde van de workshop was wie neemt het voortouw voor de verdere ontwikkeling en implementatie van het systeem? De KNAW-commissie rondt het huidige project af en komt dan met aanbevelingen. Maar dat zou niet het einde van het project moeten betekenen. De decanen hebben aangegeven dat zij de leiding willen nemen bij de ontwikkeling van tijdschriftenlijsten, de onderzoeksscholen willen daaraan meedoen. Het CWTS is ook geïnteresseerd in verdere samenwerking. De vraag is ook of er meerdere *pilots* nodig zijn? Van diverse kanten wordt aangeboden om verder te experimenteren met methodes. Er zijn ook diverse lokale initiatieven, en initiatieven in andere vakgebieden. De slotconclusie is dat er al veel gebeurt en dat er voldoende basis is om verder te gaan, maar dat het duidelijk moet worden hoe de regie is geregeld.

Rondetafel METIS-gebruikers en beleidsverantwoordelijken

Op 26 september 2011 is er bij de KNAW een bijeenkomst geweest met beleidsambtenaren en enkele onderzoekers met verantwoordelijkheid voor evaluatie. Het ging er ons in deze bijeenkomst vooral om de werkbaarheid van het systeem ter discussie te stellen. Maar ook het voorstel als geheel kwam ter sprake.

Voor wat betreft de werkbaarheid maakte men zich vooral zorgen over de aanlevering van gegevens naar METIS. De verwachting daarbij is dat vooral zal gelden voor (nieuwe) data rond de maatschappelijke relevantie. Onderzoekers klagen al vaak over overbelasting door de rapportages die ze moeten maken, en dat zal de aanlevering van nieuwe data niet makkelijker maken. De goede aanlevering van gegevens hangt over het algemeen af van het belang dat onderzoekers er in zien. In die zin is er dus een relatie met de beloningssystematiek binnen een instelling. Verschillen tussen vakgebieden spelen hierbij ook een rol. Zo ligt het in het ene vakgebied niet voor de hand om een grote mate van maatschappelijke relevantie te claimen (bv. antieke wijsbegeerte), terwijl in het andere vakgebied deze relevantie zeer evident is en kan resulteren in opdrachtonderzoek (bv. ethiek).

Bij de registratie van gegevens in METIS wordt als bijzonder probleem genoemd het feit dat sommige resultaten overlap vertonen tussen wetenschappelijke en maatschappelijke relevantie. Het moet dan mogelijk zijn om zo'n publicatie in beide categorieën op te nemen zonder dat het dubbel wordt meegeteld.

De deelnemers noemen als de meest urgente kwestie voor dit systeem dat er meer duidelijkheid komt over de wijze waarop de kwaliteit van publicaties kan worden beoordeeld, waarvoor een indeling van (wetenschappelijke) tijdschriften en uitgeverijen noodzakelijk is. Ze betreuren de mislukking in Europees verband (ERIH), maar het zou desondanks mooi zijn als het rapport in dezen voor Nederland een aanzet zou zijn. In dat perspectief beschreef iemand het rapport als 'mooi, maar de hete aardappel wordt naar voren geschoven.' Men zou willen dat de KNAW hier een initiatief nam omdat het alleen lukt als er vakgebiedoverschrijdend wordt opgetreden. Gememoreerd wordt dat bij sommige instellingen intern al wel pogingen zijn ondernomen om tot een indeling van tijdschriften te komen. Op grond daarvan schat men dat voor driekwart van de tijdschriften een indeling is te maken.

De deelnemers wijzen er op dat een systematiek voor tijdschriften en uitgeverijen een positief effect kan hebben op het gedrag van onderzoekers. Niet alleen is de verwachting dat onderzoekers meer naar de meest gewaardeerde tijdschriften trekken, de andere ('mindere') tijdschriften zullen die druk voelen en trachten te verbeteren (buitenlandse voorbeelden wijzen in die richting). Men realiseert zich dat een ordening tijd kost, maar ziet in dit systeem een kans om een begin te maken.

Uitgeverijen classificeren is moeilijker dan tijdschriften, in het bijzonder omdat uitgeverijen vaak heterogene fondsen hebben. Ook kunnen de procedures van peer review aanzienlijk verschillen. Sommige uitgeverijen hebben een duidelijke focus

(Verloren wordt als voorbeeld genoemd), anderen (bv. Waanders) publiceren zowel ansichtkaarten als serieus werk met aanzien (museumcatalogi).

Men wijst ook op de aanstaande visitatie, die geeft kansen voor nieuw beleid. Met dit rapport in de hand zouden de decanen vanaf 2012 (of als dat te vroeg komt dan de *midterm* evaluatie in 2015) helder kunnen zijn over de verwachtingen betreffende de output. De deelnemers willen meehelpen aan het totstandbrengen van een classificatie van tijdschriften, waarmee een aanbeveling van commissie Cohen en van dit rapport wordt ingelost. De deelnemers benadrukken dat ook de onderzoeksscholen hierbij moeten worden betrokken.

Gesprekspartners in de case studies

Prof. dr. S. Barbiers	Meertens Instituut, KNAW
Dr. Arianna Betti	VU, De Jonge Akademie
Prof. dr. Mineke Bosch	ICOG, Rijksuniversiteit Groningen
Drs. Klazina Botke	ICOG, Rijksuniversiteit Groningen
Prof. dr. Marcel Broersma	ICOG, Rijksuniversiteit Groningen
Drs. Liesbet Cilek	Universiteit Utrecht
Dr. Sabrina Corbellini	ICOG, Rijksuniversiteit Groningen
Drs. Heidi Disler	Rijksuniversiteit Groningen
Prof. dr. Gillis Dorleijn	ICOG, Rijksuniversiteit Groningen
Sophie Elpers, MA	Meertens Instituut, KNAW
Drs. Lisenka Fox	Bureau Onderzoek Letteren Radboud Universiteit
Frank Harbers, MA	ICOG, Rijksuniversiteit Groningen
Dr. Henk Harbers	ICOG, Rijksuniversiteit Groningen
Dr. John Helsloot	Meertens Instituut, KNAW
Dr. Ben Hermans	Meertens Instituut, KNAW
Prof. dr. Frans Hinskens	Meertens Instituut, KNAW
Dr. Margriet Hoogvliet	ICOG, Rijksuniversiteit Groningen
Prof. dr. Jan van der Horst	ICOG, Rijksuniversiteit Groningen
Drs. Linda Jansen	Universiteit van Tilburg
Dr. Willy Jongenburger	Meertens Instituut, KNAW
Dr. Willem Jongman	ICOG, Rijksuniversiteit Groningen
Drs. Erna Kas	Universiteit Utrecht
Prof. dr. Liesbeth Korthals Altes	ICOG, Rijksuniversiteit Groningen
Dr. Peter Jan Margry	Meertens Instituut, KNAW
Dr. Theo Meder	Meertens Instituut, KNAW
Konstantin Mierau, MA	ICOG, Rijksuniversiteit Groningen
Drs. Hotze Mulder	Universiteit van Amsterdam
Femke Niefhof	Meertens Instituut, KNAW
Dr. Marc van Oostendorp	Meertens Instituut, KNAW
Prof. dr. Maarten Prak	Universiteit Utrecht
Drs. Jolanda D. Riel QC	Faculteit Geesteswetenschappen, Universiteit Leiden
Dr. N. van der Sijs	Meertens Instituut, KNAW
Prof. dr. Sebastian Sobocki	ICOG, Rijksuniversiteit Groningen
Dr. Irene Stengs	Meertens Instituut, KNAW
Dr. Sandra van Voorst	ICOG, Rijksuniversiteit Groningen

BIJLAGE 2

OPDRACHT EN SAMENSTELLING VAN DE COMMISSIE

In 2009 deed de Commissie Nationaal Plan Toekomst Geesteswetenschappen (Commissie-Cohen) in haar rapport de volgende aanbeveling aan de KNAW: *‘Neem het voortouw in de ontwikkeling van een maximaal eenvoudig, helder en adequaat systeem van kwaliteitsindicatoren in de Nederlandse geesteswetenschappen.’* (pag. 45).

De KNAW heeft deze aanbeveling overgenomen, mede in het licht van de al langer lopende discussie over kwaliteitsindicatoren in de geesteswetenschappen en de behoefte die in het veld wordt gevoeld om tot een passende beoordelingsmethodiek te komen. Al eerder plaatste de KNAW kritische noten bij de toepassing van bèta-georiënteerde methoden bij kwaliteitsbeoordeling van geesteswetenschappelijk onderzoek. In 2005 publiceerde zij het rapport *Judging Research on its Merits*, waarin onder meer wordt gepleit voor ‘eigen’ meetmethoden voor de geestes- en sociale wetenschappen en voor het meewegen van de maatschappelijke waarde van deze wetenschapsgebieden.

Gezien de vraag van de Commissie-Cohen en de eerder door de KNAW geënteerde discussies over kwaliteitsbeoordeling in de sociale en geesteswetenschappen heeft het dagelijks bestuur van de KNAW in oktober 2009 de 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 bovengenoemde 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 zonodig aan te passen.
- Een eindadvies uit te brengen.

Over de eerste twee onderdelen is gerapporteerd in juni 2011 met het advies *Kwaliteitsindicatoren voor Onderzoek in de Geesteswetenschappen*. Vervolgens zijn *pilots* uitgevoerd en is op grond daarvan het bovenstaande eindadvies uitgebracht. Op persoonlijke titel werden als leden van de commissie benoemd:

- Prof. dr. Keimpe Algra, voorzitter, HL Geschiedenis van de Filosofie van de Oudheid en de Middeleeuwen, UU;
- Prof. dr. Hans Bennis, directeur Meertens Instituut KNAW, HL Taalvariatie, UvA;
- Prof. dr. Jan Willem van Henten, HL Uitlegging van het Nieuwe Testament, Vroeg-Christelijke Letterkunde en Hellenistisch-Joodse Literatuur, UvA;
- Prof. dr. Mary Kemperink, HL Moderne Nederlandse Letterkunde, RUG;
- Prof. dr. Wijnand Mijnhardt, HL Geschiedenis na de Middeleeuwen, i.h.b. Cultuur-Mentaliteits- en Ideeëngeschiedenis, UU;
- Prof. dr. Paul Wouters, directeur CWTS te Leiden, HL scientometrie, RUL (sinds januari 2011)
- Prof. dr. Paul Rutten, HL Digitale Mediastudies UL (tot 31 december 2010).
- De commissie werd ondersteund door een projectteam, bestaande uit:
- Dr. Jack Spaapen, projectleider, secretaris Raad voor Geesteswetenschappen;
- Dr. Ad Prins, pilotstudies bij Meertens en ICOG
- Dr. Mark Pen, onderzoeker (tot maart 2011)
- Dr. Koen Hilberdink, hoofd afdeling genootschap van de KNAW (tot januari 2011).