

University of Groningen

Een relationele database voor wetenschappelijke productie

Stokman, F.N.; Popping, R.; Missoorten, E.R.

Published in:
Universiteit en hogeschool

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1989

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Stokman, F. N., Popping, R., & Missoorten, E. R. (1989). Een relationele database voor wetenschappelijke productie. *Universiteit en hogeschool*, 35(6), 231-245.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Een relationele database voor wetenschappelijke produktie

door

prof. dr. F. N. Stokman, dr. R. Popping
en drs. E. R. Missoorten

PERSONALIA

De auteurs zijn verbonden aan de Vakgroep Sociologie van de Rijksuniversiteit te Groningen.

1. Uitgangspunten*

Startpunt van het onderzoeksproject om een relationele database voor de wetenschappelijke produktie uit te werken is de sterkte/zwakte analyse in het kader van de Visitatiecommissie sociologie, politicologie en bestuurskunde geweest¹⁾. De doelstelling van het project is echter aanzienlijk ruimer dan het opzetten van een database ten behoeve van het kwantificeren van wetenschappelijke produktie.

De primaire doelstelling van het project is de uitwerking van een database, waarin gegevens omtrent de wetenschappelijke produkten van de Nederlandse universiteiten in hun oorspronkelijke vorm zijn opgeslagen. Deze opslag dient zodanig te zijn dat de volgende afgeleide doelstellingen gerealiseerd kunnen worden:

1. De database dient zodanig gestructureerd te zijn dat geen beperkingen omtrent het gebruik van de database voortvloeien uit de structuur van de database. De database dient gebruikt te kunnen worden voor zowel wetenschappelijke als beleidsdoeleinden, doeleinden die de ontwerpers van de database wellicht niet kennen en voorzien. In paragraaf 5 en 6 worden enkele doeleinden genoemd waarvoor de database zou kunnen worden gebruikt, maar deze zijn slechts illustratief en niet limitatief.
2. De informatie in de database dient zodanig te kunnen worden ingevoerd, dat dit door administratieve krachten zonder kennis van de achterliggende datastructuur kan geschieden.
3. De informatie in de database dient zodanig te kunnen worden opge-

*) In de tekst zijn de opmerkingen verwerkt van de coreferent tijdens het NSAV-symposium over Scientometrie op 25 november 1988, dr. A. van Heeringen.

- vraagd dat controles door de direct betrokken onderzoekers mogelijk zijn.
4. Realisatie van de afgeleide doelstellingen 2 en 3 maakt het mogelijk dat universiteiten de database zelf gebruiken voor de produktie van hun wetenschappelijke verslagen.
 5. Realisatie van doelstelling 4 leidt tot uniformering van de verslaggeving van de wetenschappelijke produkten van de universiteiten, die noodzakelijk is om de informatie in de database te kunnen gebruiken voor wetenschappelijke en beleidsdoeleinden.
 6. Indien de universiteiten niet de database zouden gaan gebruiken voor de produktie van hun wetenschappelijke verslagen, dient de database toch de mogelijkheid te bieden om de jaarverslagen van de universiteiten op een elektronisch medium op te vragen en na bewerking in te voeren in de database.
 7. Binnen het project zal de database gevuld worden met de wetenschappelijke produktie van de Nederlandse universiteiten binnen de disciplines sociologie, politicologie en bestuurskunde om wetenschapsdynamisch onderzoek te verrichten, om de methodologie van sterkte/zwakte analyse aan te scherpen en om een door de onderzoekers zelf gecontroleerde database beschikbaar te stellen ter ondersteuning van universitaire besluitvorming omtrent een eventuele herstructurering in deze disciplines. Hiermee wordt beoogd een bijdrage te leveren aan het creëren van de voorwaarden om nieuwe ministeriële ingrepen als TVC en SKG operaties te voorkomen.

In paragraaf 2 wordt het principe van relationele databestanden uiteengezet en wordt beargumenteerd waarom juist dit type databestand geschikt is om bovenstaande doeleinden te realiseren.

Een relationele database kan worden opgevat als een bestand dat uit een (al dan niet groot) aantal rechthoekige tabellen bestaat, waarbij wel een aantal voorwaarden aan de opbouw daarvan wordt gesteld om de database relationeel te mogen noemen. Deze tabellen definiëren de relaties tussen de gegevens. In paragraaf 3 wordt dit concreet uitgewerkt voor de opslag van wetenschappelijke publikaties als voornaamste wetenschappelijk produkt. De representatie van de overige gegevens wordt slechts summier behandeld om de uiteenzetting niet nodeloos lang en technisch te maken.

In paragraaf 4 wordt ingegaan op het gebruik van het relationele databestand ten behoeve van beleidsdoeleinden. Hierbij wordt ingegaan op de sterkte/zwakte analyse van de Visitatiecommissie sociologie, politicologie en bestuurskunde en op de extra mogelijkheden die de relationele database heeft. Dit wordt mede geplaatst in het kader van internationale ontwikkelingen rond het meten van wetenschappelijke produkties en kwaliteit. In paragraaf 5 worden illustraties gegeven voor het gebruik van de database ten behoeve van wetenschapsdynamisch onderzoek. Tenslotte worden in paragraaf 6 enkele algemene kanttekeningen geplaatst om een onjuist gebruik van de database voor beleidsdoeleinden te voorkomen.

2. De voordelen van relationele databestanden om gegevens over wetenschappelijke produkten op te slaan

In de sterkte/zwakte-analyse van de visitatiecommissie is, daar waar het ging om de publikaties, vooral gebruik gemaakt van tellingen. Gegeven de doeleinden die de commissie zich had gesteld en de tijdsspanne waarin het een en ander diende te geschieden was dat een begrijpelijke keuze voor een bepaalde manier van gegevensrepresentatie.

In de rapportage van deze analyse werd echter tevens de wens geuit om verdere analyses te verrichten: analyses op verschillende aggregatieniveau's en analyses waarin deze niveau's aan elkaar kunnen worden gerelateerd. Die wens bracht de beperkingen van het gebruik van tellingen aan het licht. Voor het uitvoeren van bovengenoemde verdere analyses is de gegevensrepresentatie door middel van tellingen ongeschikt.

De twee belangrijkste redenen hiervoor zijn:

- Er kan niet voldoende gedifferentieerd worden naar individuele attributen of kenmerken van een bepaald gegevenstype, bijvoorbeeld publikaties.
- Gegevensrepresentatie door middel van tellingen maskeert het netwerk van relaties tussen de verschillende gegevens.

Wil men bijvoorbeeld controles uitvoeren naar de juiste categorisering van een bepaalde publikatie, dan is het noodzakelijk om individuele publikaties te kunnen identificeren en het kenmerk te kunnen bepalen waarop de indeling heeft plaatsgevonden. Dat is alleen mogelijk wanneer men beschikt over een gegevensrepresentatie waarin iedere publikatie individueel is gespecificeerd met de daarbijbehorende attributen. Bij representatie van publikaties door middel van tellingen kan immers niet meer onderscheiden worden naar individuele publikaties, laat staan dat er nog gekeken kan worden naar een bepaald kenmerk van een bepaalde publikatie.

Ook om andere redenen kan het noodzakelijk zijn om te kunnen bepalen wat de waarde is van een bepaald attribuut van een publikatie. Als men over een telling beschikt die gemaakt is over een aantal publikaties met betrekking tot een bepaald kenmerk, dan is het niet meer mogelijk om, uitgaande van die telling, te kunnen bepalen wat het resultaat zou zijn indien men een andere indeling binnen de verschillende waarden van dat kenmerk zou hanteren of een ander kenmerk als criterium zou gebruiken.

Zo is het bijvoorbeeld mogelijk dat men een analyse wil verrichten naar de mogelijke voorkeur van verschillende onderzoeksgroepen voor het publiceren in bepaalde media. Een dergelijke wens zou mogelijkerwijs kunnen voortkomen vanuit de behoefte het al of niet bestaan van scholen binnen de Nederlandse sociologie nader te bestuderen. Daarvoor is het echter noodzakelijk om enerzijds te kunnen bepalen in welk medium een bepaalde publikatie is gepubliceerd en tevens te kunnen aanwijzen welke auteurs, en daarmee onderzoeksgroepen, verantwoordelijk zijn voor die bepaalde publikatie.

Daarmee komen we aan bij het tweede bovengenoemde bezwaar, namelijk dat de relaties tussen verschillende gegevenstypen worden gemaskeerd. De relaties publikaties/media en publikaties/auteurs zijn twee voorbeelden van relaties die bij het hanteren van tellingen niet onderscheiden kunnen worden. Het mag duidelijk zijn dat, wil men tegemoetkomen aan de bovengestelde wensen, er gekozen moet worden voor een andere manier van gegevensrepresentatie. Voor de wetenschappelijke publikaties ligt het voor de hand de publikaties zelf te representeren. Dat wil zeggen dat elke publikatie individueel met inbegrip van alle kenmerken als element in de dataset wordt vertegenwoordigd. Voor analyses waar de onderzoekseenheid de publikatie is en analyses die zich richten op een van de attributen van de publikatie is het immers noodzakelijk dat dat kenmerk aanwezig is met de waarden waaruit dat kenmerk bestaat in de onderzoeksdataset. Bij de manier van gegevensopslag en beheer binnen het onderzoek is gekozen voor een geautomatiseerd gegevensbestand. Het gebruik daarvan biedt verschillende voordelen. Naast een aanmerkelijke verbetering in de toegangssnelheid waarmee de gebruiker de dataset kan benaderen en de vele mogelijkheden op het gebied van gegevensinvoer, biedt een geautomatiseerd gegevensbestand (ook wel genoemd: DataBaseManagementSysteem - hierna DBMS -) belangrijke voorzieningen met betrekking tot de organisatie van de gegevensstructuur en de specificatie van de relaties tussen verschillende gegevens in het gegevensbestand. Deze voordelen worden met name geboden in een op het relationeel model gebaseerd DBMS²). Eén van de voordelen die boven traditionele gegevensmodellen, zoals het netwerkmodel en het hiërarchische model, geboden wordt, betreft de manier waarop relaties tussen gegevens gespecificeerd kunnen worden. Binnen de meer traditionele manier van gegevensmodellering is het namelijk noodzakelijk dat relaties tussen de gegevens en de gehele gegevensstructuur op voorhand gespecificeerd worden. Als het gegevensmodel eenmaal op de database geïmplementeerd is, kan achteraf geen wijziging of toevoeging aan de gegevensstructuur aangebracht worden, zonder het gehele gegevensmodel aan te moeten passen aan de wijziging en vervolgens opnieuw te moeten implementeren.

Het relationele model garandeert daarentegen een principiële onafhankelijkheid tussen de feitelijke gegevens en de structuur van die gegevens. Die onafhankelijkheid vindt zijn gedaante in de aparte opslag van gegevens en structuur. Onder gegevensstructuur wordt hier onder andere verstaan de fysieke plaats van informatie op het opslagmedium, de manier waarop dat is georganiseerd en de relaties die tussen de gegevens gelegd kunnen worden. Om dat toe te lichten nemen we hier de relatie persoon en vakgroep. Een persoon heeft als kenmerk dat deze tot een bepaalde vakgroep behoort. Verandert na verloop van tijd deze persoon van vakgroep, dan dient het geheel opnieuw opgeslagen te worden. Bij het relationele model echter, wordt de informatie die rechtstreeks op de persoon van toepassing is, bijvoorbeeld naam en voorletters, gescheiden opgeslagen van de informatie met betrekking tot de relatie met de vakgroep. In **figuur 1** wordt dat zichtbaar gemaakt.

Figuur 1. Een voorbeeld van tabellen in een relationele database.

personen

persoonsnummer	naam	voorletters
0001	jansen	j

functies

persoonsnummer	vakgroepnummer	begindatum	einddatum
0001	0001	01/01/86	31/12/87
0001	0002	01/01/88

vakgroepen

vakgroepnummer	vakgroepnaam
0001	methoden en technieken
0002	arbeid en organisatie

In deze figuur wordt de basis waarop het relationele model stoelt tot uitdrukking gebracht: De informatie in een relationele database is opgeslagen in tabelvorm. In de figuur is een verandering aangebracht in de relatie die jansen heeft met een bepaalde vakgroep, terwijl de basisgegevens met betrekking tot de persoonsgegevens en de vakgroepgegevens ongewijzigd zijn gebleven.

Elke tabel in een relationele database dient een primaire sleutel te hebben. Deze bestaat uit één of meer kolommen uit een tabel die de eigenschap hebben dat ze een rij in die tabel uniek identificeren.

Een kolom waarop de relatie tussen twee tabellen wordt gelegd noemen we een refererende sleutel. Deze hoeft niet uniek te zijn. De relatie tussen de tabellen personen en functies kan bijvoorbeeld gelegd worden door de kolom persoonsnummer, die in beide tabellen aanwezig is, als refererende sleutel te nemen. Op soortgelijke manier kan een verbinding worden gelegd tussen

de tabellen functies en vakgroepen via de kolom vakgroepnummer. Er kan zelfs een relatie worden gedefinieerd tussen de tabellen personen en vakgroepen. Deze hebben op zichzelf geen kolommen gemeen maar de verbinding kan hier via de tabel functies tot stand worden gebracht.

In de volgende paragraaf wordt een relationeel gegevensbestand voor wetenschappelijke publikaties uitgewerkt. De informatie waarmee het bestand gevuld wordt, is opgenomen in de wetenschappelijke jaarverslagen die jaarlijks door universiteiten en hogescholen worden uitgegeven.

Voordat dieper op het publikatiedatabestand wordt ingegaan, moet er eerst iets gezegd worden over die jaarverslagen.

Aan een databaseontwerp gaat meestal een analyse met betrekking tot de vorm en aard van de te gebruiken gegevens en gegevensdragers vooraf. Deels heeft zo'n analyse ten doel te komen tot een zodanige stroomlijning in de aanvoer van de gegevens, dat bij invoer van de gegevens geen onnodige extra werkzaamheden buiten de invoer zelf nodig zijn. Zulke extra werkzaamheden, zoals het coderen of combineren van gegevens vanaf verschillende gegevensdragers, leiden slechts tot fouten en vertragingen. De ideale gegevensstroom heeft als kenmerken onder andere uniformiteit in de vorm waarin de gegevens aangeleverd worden en duidelijkheid qua aard. Helaas zijn deze kenmerken op het samenstel van de verschillende wetenschappelijke jaarverslagen niet van toepassing. Hierin wordt nog een reden gevonden om te pleiten voor het gebruik van een geautomatiseerd gegevensbestand bij opslag van gegevens die opgenomen zijn in die jaarverslagen. Dat geldt eens te meer als ook overeenstemming kan worden bereikt over een bepaalde standaardisatie in de programmatuur.

3. Uitwerking van een relationeel databestand voor wetenschappelijke publikaties

Bij het modelleren van de gegevensstructuur rond de publikaties duiken er een paar problemen op. De gegevensstructuur van een publikatie is te ingewikkeld om ondergebracht te kunnen worden in één gegevenstype.

In essentie hebben we te maken met drie verschillende gegevenstypen:

- Boeken, waaronder readers en dissertaties;
- Tijdschriftartikelen;
- Readerbijdragen.

De drie gegevenstypen verschillen in de informatie die in referenties over het medium gegeven worden: bij boeken behoren uitgeversgegevens, bij tijdschriftartikelen behoren gegevens met betrekking tot de tijdschriften waarin ze opgenomen zijn en bij readerbijdragen horen gegevens over de reader. Bij traditionele gegevensmodellering zou in dit geval gekozen worden voor een indeling in drie verschillende tabellen: één voor de boeken, één voor de

tijdschriftartikelen en één voor de readerbijdragen. In de drie tabellen zijn dan tevens de gegevens opgenomen met betrekking tot de media. Dat kan betekenen dat de media- en auteursgegevens in sommige gevallen meerdere malen in de database opgenomen moeten worden. In het relationele model is het mogelijk om de basisgegevens van de publikaties onder te brengen in één tabel, namelijk de tabel publikaties. In die tabel is een attribuut opgenomen, publikatiesoort, dat de waarde boek, (reader)bijdrage of (tijdschrift)artikel kan aannemen. Tevens is een mediumnummer als kenmerk opgenomen. Het veld publikatiesoort bepaalt de betekenis van mediumnummer. Het mediumnummer komt ook voor in de tabellen waar de informatie is opgeslagen betreffende de media waarin de publikatie is gepubliceerd: de tabellen tijdschriften, readers en uitgevers. Het kenmerk mediumnummer vormt een refererende sleutel naar die tabellen. In de tabel tijdschriften wordt de verbinding gelegd via het attribuut tijdschriftnummer en in de tabel uitgevers wordt de verbinding tot stand gebracht via het uitgeversattribuut uitgeversnummer. Bij publikaties die geklassificeerd worden als readerbijdrage wordt de relatie tussen bijdrage en reader gelegd in de tabel readers die bestaat uit de attributen: bijdragenummer en readernummer. De verbinding met de tabel uitgever verloopt hier niet rechtstreeks maar via de tabellen readers en publikaties zelf. In de tabel publikaties zijn namelijk ook die readers opgenomen met een verwijzing naar de uitgever via het mediumnummer.

Samenvattend: heeft het veld publikatiesoort als inhoud de waarde „boek”, dan verwijst het veld mediumnummer naar het veld uitgeversnummer in de tabel uitgevers. Als de inhoud „bijdrage” is, verwijst mediumnummer naar een publikatienummer van een ander record in dezelfde tabel, namelijk de reader die eveneens als publikatie in de tabel is opgenomen. Bij de waarde „artikel” vervult het veld mediumnummer de functie van refererende sleutel vanuit de tabel tijdschriftmedia.

Informatie omtrent de relatie tussen publikaties en auteurs wordt opgeslagen in de tabel auteurs. De attributen persoonsnummer en publikatienummer verwijzen naar de gelijknamige kolommen in respectievelijk de tabellen personen en publikaties.

Beschouwen we de tabel persoonsgegevens als behorende tot het gegevenstype publikaties, dan kunnen we stellen dat de publikatiegegevens in zes tabellen zijn opgenomen.

Dat dat een eenvoudiger manier van gegevensmodellering is, ligt zo voorgesteld niet direct voor de hand.

Men moet daarbij echter in het oog houden dat essentiële gegevens van de publikaties, zoals titels, taal, jaar van uitgifte en dergelijke in één tabel zijn ondergebracht en dat bijvoorbeeld uitgevers en tijdschriften slechts één keer opgeslagen hoeven te worden en daarna altijd beschikbaar zijn voor nieuwe publikaties.

Hieronder zijn de zes tabellen gegeven met de bijbehorende attributen. Daarbij zijn de kolommen die tot de primaire sleutel behoren cursief weergegeven.

Publikaties:

<i>publikatienummer</i>	-	identificatienummer van de publikatie.
titel	-	de titel.
ondertitel	-	de ondertitel.
inhoud	-	wetenschapsgebied (NABS).
soort	-	soort publikatie (boek, tijdschriftartikel of readerbijdrage).
taal	-	taal waarin de publikatie geschreven is.
jaar	-	jaar van uitgave.
beginpagina	-	beginpagina.
eindpagina	-	eindpagina.
auteurvolgorde	-	of er een volgorde bestaat, anders dan een alfabetische, in de manier waarop de auteurs vermeld zijn.
mediumnummer	-	identificatienummer van het medium waarin de publikatie is gepubliceerd.

Uitgevers:

<i>uitgeversnummer</i>	-	identificatienummer van de uitgever.
uitgeversnaam	-	naam van de uitgever.
plaats	-	plaats waar de uitgever gevestigd is.

Readers:

<i>readerbijdragenummer</i>	-	publikatienummer van de readerbijdrage.
<i>readernummer</i>	-	publikatienummer van de reader.

Auteurs:

<i>persoonsnummer</i>	-	persoonsnummer van de auteur.
<i>publikatienummer</i>	-	identificatienummer van de publikatie.
auteursoort	-	hoedanigheid van het auteurschap (redacteur, auteur e.d.).
volgnummer	-	volgnummer van de auteur in de auteurslijst van de publikatie.

Tijdschriften:

<i>tijdschriftnummer</i>	-	identificatienummer van het tijdschrift.
naam	-	naam van het tijdschrift.

Personen:

<i>persoonsnummer</i>	-	identificatiecode van de persoon.
naam	-	de achternaam.
voorletters	-	de voorletters.
voorvoegsels	-	de voorvoegsels.

Hiermee zijn de essentiële gegevens omtrent wetenschappelijke publikaties opgenomen in een relationeel databestand. Voor details betreffende de applicatie voor de invoer van de gegevens wordt verwezen naar STOKMAN et al. (1988).

4. Gebruik van de gegevens ten behoeve van beleidsdoeleinden

In deze paragraaf gaat het om de mogelijkheden van relationele databases om gegevens - op overzichtelijke wijze en aangepast aan de doelstellingen van de gebruiker - uit de database te extraheren ten behoeve van beleidsdoeleinden. Relationele database systemen zijn hiervoor uitermate geschikt, omdat zij de mogelijkheid bieden speciale applicaties te schrijven voor regelmatig terugkomende extracties van gegevens uit het gegevensbestand. Hierbij wordt gebruik gemaakt van een krachtige zoek- en vraagtaal, SQL genoemd (EILERS e.a. 1986).

Ons gaat het er hier meer om de mogelijkheden van het gebruik van de database voor beleidsdoeleinden te illustreren dan om een technische uiteenzetting te geven over de wijze waarop dit kan worden gerealiseerd.

Hierboven hebben wij reeds de mogelijkheid genoemd een applicatie te schrijven om rechtstreeks vanuit de database wetenschappelijke verslagen te genereren. In paragraaf 4.1 wordt het gebruik van de database voor de meting van kwaliteit en impact van onderzoeksgroepen behandeld. In paragraaf 4.2 worden vervolgens mogelijkheden voor beleidsdoeleinden besproken, waaronder het gebruik van de database voor een bredere sterkte/zwakte analyse.

4.1 Meting van kwaliteit en impact van onderzoeksgroepen

Massale opslag van bibliografische gegevens van onderzoeksgroepen kan primair gebruikt worden ten behoeve van onderzoek naar kwaliteit en impact van onderzoeksgroepen met behulp van „science indicators”.

Grosso modo vindt men in de literatuur twee „science indicators” terug: een (gewogen) som van verschenen publikaties en een (gewogen) som van citaties.

De discussie rond validiteit van tellingen van publikaties als indicator van wetenschappelijke output concentreert zich rond de problemen bij toewijzing van kenmerken van publikaties bij co-auteurschap en bij weging van het publikatiemedium naar kwaliteit en prestige.

Bij gebruikmaking van wetenschappelijke indicatoren is het onmogelijk elke publikatie naar kwaliteit te meten³). Vandaar dat weging doorgaans naar twee criteria plaatsvindt: naar de omvang en naar de kwaliteit van het medium waarin de publikatie is opgenomen. Deze laatste wordt veelal bepaald op basis van het aantal keren dat publikaties uit dat medium worden geciteerd. Problemen bij de weging van auteurs komen doorgaans voort uit de starheid van de gebruikte analytische databases waarbij de analyse-eenheid publikatie niet goed is te scheiden van de analyse-eenheid auteur. Lindsey (1980) heeft een overzicht gegeven van de wijze waarop telling van publikaties in 40 verschillende studies heeft plaatsgevonden. Er wordt op drie manieren geteld:

- alle kenmerken worden bij elke auteur eenmaal geteld;
- alle kenmerken worden gewogen geteld naar het aantal auteurs;
- alle kenmerken worden alleen geteld voor de eerste auteur.

Lindsey pleit voor de tweede manier van tellen. Hij laat zien dat bij de andere manieren nogal wat meetfouten ontstaan. Bij de derde manier is het ook zo dat er met een steekproef wordt gewerkt en het is evident dat deze niet representatief is.

Het door Lindsey bepleite alternatief is, zoals hiervoor is duidelijk gemaakt, via een relationele database eenvoudig te realiseren.

Voor een zinvolle telling van publikaties is het in elk geval nodig dat wetenschappelijke verslagen volledig zijn en dat de categorie waarin de wetenschappelijke publikaties mogen worden opgenomen eenduidig is met betrekking tot de vraag welke publikaties als wetenschappelijk mogen worden aangemerkt. Dit laatste is niet per publikatie te bepalen. Dergelijke criteria kunnen daarom het beste geformuleerd worden in termen van het medium waarin de publikatie is opgenomen zoals al eerder is gedaan met betrekking tot de kwaliteit van de publikaties.

We gaan hier niet in op problemen bij citaties. Voor een eerste overzicht wordt verwezen naar Martin en Irvine (1983), King (1987) en de opmerkingen die hierover door Van Loon c.s. (1988) zijn gemaakt.

In het databestand zijn (nog) geen tabellen opgenomen die kwaliteit van de media beogen te meten, zoals het prestige van een tijdschrift. Dergelijke gegevens zijn eenvoudig aan het databestand toe te voegen via extra tabellen waarbij de mogelijkheid bestaat het prestige in de tijd te laten variëren.

De wijze waarop de publikaties in het databestand zijn opgenomen laat alle mogelijkheden toe tot kwantificatie - met weging naar omvang en prestige van het medium - per persoon of onderzoeksgroep, waarbij zeer vergaande analyses op basis van de door de onderzoekers gekozen media mogelijk zijn. Wanneer citaties zo aan het bestand worden toegevoegd dat een relatie wordt gelegd tussen de geciteerde publikatie en de publikatie waarin de cita-

tie plaats vindt, is het mogelijk citaties zowel per publikatie, per persoon als per onderzoeksgroep te sommeren, waarbij desgewenst gewogen kan worden naar het prestige van het medium waarin de citatie plaats vindt. Omdat alle bibliografische gegevens zelf in het bestand zijn opgenomen en alle relaties tussen publikaties enerzijds en personen en onderzoeksgroepen anderzijds uit het bestand kunnen worden afgeleid - zonder dat zij daarin expliciet zijn gedefinieerd - worden a priori geen beperkingen vanuit de datastructuur opgelegd op grond van bijvoorbeeld de huidige inzichten bij citatieanalyse.

4.2 Gebruik van de database voor beleidsdoeleinden

Van groot belang voor het beleid is de vraag wat de structurele voorwaarden zijn om tot een sterke onderzoeksgroep te komen en welke kenmerken positief met sterkte of zwakte van onderzoek correleren. Aangezien de overheid zich wil gaan concentreren op haar voorwaarden-scheppende taken, is deze vraag naar de condities voor kwaliteit van groot belang. Juist omdat de database geen beperkingen oplegt aan het soort vragen dat aan de database gesteld kan worden, is zij uitermate geschikt voor dergelijk beleidsgericht onderzoek. Zo is het ministeriële beleid nogal gericht op vergroting van de onderzoekseenheden om tot een grotere efficiëntie te komen. Cramer (1988) laat zien dat dit althans voor de economische wetenschap in Nederland iedere grond mist. Hij laat verder zien dat schommelingen in kwaliteit sterk lijken samen te hangen met mobiliteit van personen en concludeert: „Wetenschappelijke produktiviteit is geen eigenschap van vakgroepen, faculteiten of andere instituten, maar van mensen die er werken. Als zij van vakgroep veranderen nemen zij hun produktiviteit mee” (CRAMER, 1988, 12).

De mogelijkheden van een relationele database om van analyse-eenheid te veranderen, maakt haar bij uitstek geschikt om de vraag te beantwoorden naar het primaat van de eenheid die de kwaliteit draagt. Bij een dergelijke analyse is het mogelijk effecten van institutionele factoren op de produktiviteit van individuele onderzoekers respectievelijk onderzoeksgroepen te bepalen en af te zetten tegen effecten van persoonlijke kenmerken van onderzoekers daarop. Een voorbeeld van een dergelijke analyse, maar dan slechts op het niveau van de individuele onderzoeker, is het onderzoek van Van Heeringen en Dijkwel (1987a en 1987b).

Met betrekking tot de sterkte/zwakte analyse van onderzoeksgroepen merken wij op dat ook andere kenmerken van onderzoeksgroepen in de analyse dienen te worden betrokken dan publikaties en citaties. De Visitatiecommissie sociologie, politicologie en bestuurskunde ging in het totaal uit van acht criteria. Om de kwaliteitsscore minder gevoelig te maken voor kleine verschillen, werd gebruik gemaakt van kwartielscores. Voor de acht criteria en resultaten van de sterkte/zwakte analyse voor vakgroepen en studierichtingen wordt verwezen naar Van Loon c.s. (1988).

In de discussie zijn zowel de keuze van de criteria als hun onderlinge weging via de ongewogen sommatie van de kwartielscores aangevallen. Mede naar aanleiding van deze discussie geven Van Loon en Stokman 11 punten aan waarop in hun ogen de gevolgde methode van sterkte/zwakte analyse kan worden verbeterd (VAN LOON en STOKMAN, 1988, 124-125). Wij onderschrijven deze desiderata. Met uitzondering van twee desiderata die in de beleidssfeer liggen, zijn ALLE desiderata met de voorgestelde database te realiseren, vooral wanneer deze inderdaad zelf gebruikt zou gaan worden voor de produktie van de wetenschappelijke verslagen van de universiteiten. Voorwaarde hiertoe is natuurlijk dat ook de andere gegevens dan die over publikaties in soortgelijke tabellen aan de database zijn toegevoegd. Dit is in ons ontwerp inderdaad gebeurd.

5. Gebruik ten behoeve van wetenschappelijke doeleinden

Stond in paragraaf 4 de produktiviteit van onderzoekers en onderzoeksgroepen - en daarvan afgeleid de kwaliteit en impact - centraal, in deze paragraaf worden kort enkele voorbeelden gegeven van het gebruik van de database voor meer wetenschapsdynamisch onderzoek.

Een eerste voorbeeld betreft de vraag naar samenwerkingspatronen tussen onderzoekers en onderzoeksgroepen op basis van co-auteurschap. Een longitudinaal onderzoek daarnaar is juist nu uitermate interessant wegens de fundamentele verbeteringen in de communicatiemogelijkheden ten gevolge van de elektronische netwerken, zowel nationaal als internationaal. Deze vraagstelling is nauw verwant aan de vraagstelling in paragraaf 4.2, zij het dat nu de analyse niet gericht is op institutionele effecten op de produktiviteit, maar op de institutionele effecten op frequentie en het patroon van samenwerking tussen onderzoekers.

Een tweede voorbeeld betreft de vraag in hoeverre er sprake is van schoolvorming binnen diverse disciplines. Dergelijke analyses worden veelal verricht op basis van een analyse van het citatienetwerk. Een voorbeeld daarvan is het onderzoek van Hagendijk en Prins (1984) naar de aard van de wetenschappelijke produktie van de beoefening van de sociologie en de receptie van het werk van sociologen in de belangrijkste Nederlandstalige sociologische tijdschriften. Vanuit wetenschapsdynamisch gezichtspunt gezien is het interessant na te gaan of een dergelijke schoolvorming ook blijkt uit andere relaties die uit publikaties zijn af te leiden. Men kan hierbij enerzijds denken aan een veel sterkere relatie dan de citatie, namelijk het co-auteurschap, en anderzijds aan een zwakkere relatie dan de citatie, namelijk de keuze van de media, waarin het onderzoek wordt gepubliceerd.

Een derde voorbeeld van het belang van de database voor meer algemeen wetenschapsdynamisch onderzoek is gelegen in het feit dat de database in

principe alle wetenschappelijke publikaties vanuit de Nederlandse universiteiten zal bevatten. Hiermee wordt de database uitermate interessant als raster voor steekproeftrekking, waardoor bijvoorbeeld analyses van de inhoud van publikaties - bij een juiste procedure van steekproeftrekking - generaliseerbaar worden naar de populatie van alle wetenschappelijke publikaties vanuit de Nederlandse universiteiten of deelpopulaties daarbinnen. Onderzoek op het gebied van wetenschapssociologie waarbij gebruik wordt gemaakt van referenties en citaties is niet nieuw. Door De Solla Price (1963) is er al veel aandacht aan besteed. Voor een recent overzicht wordt verwezen naar King (1987).

Tenslotte is het in beginsel mogelijk de database geschikt te maken voor het verschaffen van informatie over de inhoud van onderzoek. Te denken valt aan (co-)woord analyse van titels, het opnemen van trefwoorden en mogelijk abstracts.

6. Slotopmerkingen

Een relationele database voor wetenschappelijke output is besproken. Als universiteiten zouden overwegen een dergelijke database te gebruiken voor de produktie van hun wetenschappelijk verslag, dan zou de database beperkt kunnen blijven tot de tabellen zoals hierboven gepresenteerd, met toevoeging van enkele tabellen voor omvang van onderzoekscapaciteit. Deze database zou dan voor specifiekere doeleinden ten behoeve van beleidsgericht en wetenschappelijk onderzoek beschikbaar moeten worden gesteld, waarbij onderzoekers de database kunnen uitbreiden met gegevens die voor dat onderzoek additioneel nodig zijn. Een dergelijke uitgebreide database wordt op dit moment gevuld voor de disciplines sociologie, politicologie en bestuurskunde in het kader van ons eigen project. Wanneer het algemeen bekend wordt dat de wetenschappelijke jaarverslagen hiervoor gebruikt zullen gaan worden, zal alleen al die voorkennis bij de onderzoekers tot een grotere nauwkeurigheid in hun opgave leiden. De tendens om dan ook allerlei onofficiële publikaties op te nemen zal hiermee alleen maar toenemen en dient derhalve met stringente richtlijnen te worden teniet gedaan. Ook zal de neiging verder toenemen om vanuit gevestigde posities het co-auteurschap van publikaties op te eisen, zelfs indien de bijdrage verwaarloosbaar, klein of nihil is. Hofstee (1988) maakt zich hier terecht bezorgd over en geeft hieromtrent zeer weloverwogen richtlijnen.

Gebruik voor beleidsgericht onderzoek vereist naast de gegevens over de output betrouwbare gegevens omtrent de inzet van middelen. Het is algemeen bekend dat ook aan deze input-kant de wetenschappelijke jaarverslagen niet altijd even betrouwbaar zijn. Ook hieromtrent dienen duidelijke afspraken gemaakt te worden en dienen controles uitgevoerd te worden of de inzet niet geminimaliseerd wordt om een zo groot mogelijke relatieve output

te verkrijgen. Dergelijke controles vereisen een vergelijking van de inzet voor onderzoek met die voor onderwijs en overige taken. Valide gegevens zijn een noodzakelijke voorwaarde voor relevant beleidsgericht onderzoek. Zij zijn natuurlijk geen voldoende voorwaarde voor een verstandig gebruik van de gegevens bij de beleidsbepaling. Beleidsbepalers dienen zich te blijven realiseren dat kwantitatieve gegevens uitsluitend als aanvullende evidentie gebruikt kunnen worden om kwalitatieve beoordelingen door deskundigen te ondersteunen, zij kunnen deze nooit vervangen. Realisering van ons voorstel zou echter een unieke kans betekenen de betrouwbaarheid en stabiliteit van allerlei kwantitatieve indicatoren op langere termijn te onderzoeken. Hiervoor is wel een terughoudend gebruik van de gegevens voor beleid op korte termijn noodzakelijk.

Noten

- 1) *Rapport Visitatiecommissie*, 1987; voor een nadere rechtvaardiging en discussie zie VAN LOON e.a., 1988 in *Acta Politica*, de reacties daarop van DAALDER, HUZEN, KÖBBEN en VAN SCHENDELEN en het weerwoord van VAN LOON en STOKMAN in hetzelfde nummer van *Acta Politica* (23, 1, januari 1988).
Waar in deze bijdrage zonder nadere specificatie sprake is van „de visitatiecommissie”, is steeds de Visitatiecommissie voor sociologie, politicologie en bestuurskunde bedoeld. Deze commissie werd, in het kader van de SKG-operatie, in november 1987 door de minister van onderwijs en wetenschappen ingesteld en bracht in februari 1988 haar bovengenoemde rapport uit.
- 2) Voor een verdere behandeling van de theoretische achtergronden van het relationele datamodel wordt verwezen naar: DATE, 1985, JANSSENS, 1987 en VAN DER LANS, 1987.
- 3) Moed et al. onderscheiden drie vormen van kwaliteit die elkaar kunnen overlappen: cognitieve kwaliteit, waarbij het gaat om het wetenschappelijk belang van hetgeen in het artikel vermeld staat; methodologische kwaliteit, dit verwijst naar de accuraatheid van de gebruikte methoden en technieken; en esthetische kwaliteit, waarbij gelet op op de vorm waarin een en ander is gepresenteerd (MOED et al. 1985, 134).

Literatuur

- CRAMER, J. S., De produktiviteit van economische vakgroepen nader bezien. In: *Kwantitatieve Methoden*, 9 (1988) 29, 5-12.
- DAALDER, H., Beleid, visitatie en kwaliteitsbeoordeling. In: *Acta Politica*, 23 (1988) 61-100.
- DATE, C. J., *Database, een inleiding*. Den Haag: Academic Service, 1985.
- DE SOLLA PRICE, D. J., *Little Science, Big Science*. New York, Columbia University Press, 1963.
- EILERS, H. B., W. F. JANSSEN & H. H. J. DE VOLDER, *SQL in de praktijk*. Den Haag: Academic Service, 1986.
- HAGENDIJK, R. P. & A. A. M. PRINS, *Referenties en Reverences. Onzekerheid, afhanke-*

- lijkheid en citeernetwerken in de Nederlandse sociologie. In: *Mens en Maatschappij*, 59 (1984) 226-250.
- HEERINGEN, A. VAN & P. A. DIJKWEL, The relationships between age, mobility and scientific productivity. Part I. Effect of Mobility on Productivity. In: *Scientometrics*, 11 (1987a) 267-280.
- HEERINGEN, A. VAN & P. A. DIJKWEL, The relationships between age, mobility and scientific productivity. Part II. Effect of Age on Productivity. In: *Scientometrics*, 11 (1987b) 281-293.
- HOFSTEE, W. K. B., Auteursvolgorden bij wetenschappelijke artikelen. In: *U&H Tijdschrift voor wetenschappelijk onderwijs*, 5 (1987-1988) 211-213.
- HUZEN, Th. P., Schaamte, waar is uw blos? Kritische kanttekeningen bij: Sterkte/zwakte-analyse sociologie, politicologie en bestuurskunde. In: *Acta Politica*, 23 (1988) 101-108.
- JANSSENS, H., *Principes van relationele databanken*. Deventer: Kluwer, 1987.
- KING, J., A Review of Bibliometric and Other Science Indicators and their Role in Research Evaluation. In: *Journal of Information Science*, 13 (1987) 261-276.
- KÖBBEN, A. J. F., Commentaar. In: *Acta Politica*, 23 (1988) 109.
- LANS, R. F. VAN DER, *Het relationele model*. Den Haag: Academic Service, 1987.
- LINDSEY, D., Production and Citation Measures in the Sociology of Science: The Problem of Multiple Authorship. In: *Social Studies of Science*, 10 (1980) 145-162.
- LOON, P. J. J. M. VAN & F. N. STOKMAN, Reactie op Daalder, Huzen en Köbben. In: *Acta Politica*, 23 (1988) 122-134.
- LOON, P. J. J. M. VAN, F. N. STOKMAN, C. VAN DER HUUCHT & P. MOORER, Sterkte/zwakte-analyse sociologie, politicologie en bestuurskunde. In: *Acta Politica*, 23 (1988) 21-60.
- MARTIN, B. R. & J. IRVINE, Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy. In: *Research Policy*, 12 (1983) 61-90.
- MOED, H. F., W. J. M. BURGER, J. G. FRANKFORT & A. F. J. VAN RAAN, The Use of Bibliometric Data for the Measurement of University Research Performance. In: *Research Policy*, 14 (1985) 131-149.
- Rapport van de visitatiecommissie voor sociologie, politicologie en bestuurskunde* (R. S. J. M. VAN STRAATEN, P. J. J. M. VAN LOON, G. A. KOOY en F. N. STOKMAN) 1987.
- SCHENDELEN, M. P. C. M. VAN, Een Talleyrands eerherstel voor de visitatiecommissie. In: *Acta Politica*, 23 (1988) 111-121.
- STOKMAN, F. N., R. POPPING en E. MISSOORTEN, *Een relationele database voor wetenschappelijke produktie*, ICS-working paper 3, Groningen, 1988.