KM 28(1988)
pag 31 - 50

ED AERTS EN TON DUFFHUES[1]

RELATIONELE DATABASES IN SOCIAAL-WETENSCHAPPELIJK ONDERZOEK: EEN PLEIDOOI

SAMENVATTING

In dit artikel laten we zien dat het relationele model van gegevens-opslag van groot belang is in sociaal-wetenschappelijk en historisch onderzoek[2]. Vergelijking met de conventionele manier van gegevens-opslag in de rechthoekige datamatrix leert dat het relationele model voordelen biedt, en zeker aantrekkingskracht bezit voor onderzoek dat niet strict is opgezet volgens het klassieke survey. In het eerste deel van het artikel wordt een vergelijking tussen het conventionele en relationele model uitgewerkt. In het tweede deel wordt aangetoond waarom zelfs in de sociaal-netwerkanalyse, waarin men al beschikt over een model op basis van relaties, relationele databases toch zinvol zijn. Mede op basis hiervan wordt geconcludeerd dat onderzoekers er verstandig aan doen om bij de opslag van gegevens toepassing van het relationele model te overwegen. In het verlengde hiervan ligt het pleidooi om statistische software aan te passen aan een dergelijke wijze van data-opslag.

INLEIDING

Maar al te vaak stuiten we in sociaal-wetenschappelijk onderzoek tijdens het verzamelen van gegevens op de beperkingen van de opslag in een rechthoekige matrix. Met name als een respondent meer antwoorden kan geven op één vraag, of als informatie verzameld wordt uit verschillende bronnen, worden meestal vooraf een aantal keuzes gemaakt die beperkingen opleggen. Bij veel multi-respons ligt de oplossing vaak nog voor de hand: eerst bepalen hoeveel mogelijke keuzes er zijn en daarna in de matrix die keuzes aangeven. Menig onderzoek(st)er zal echter de bedenkingen herkennen die opborrelen na de data-opslag: 'had ik maar een andere indeling van mijn multi-respons-kategorieën gemaakt'. De beperkingen worden helaas voor lief genomen, want men

[1]Beide werken voor de Vakgroep Methoden van Maatschappijwetenschappen en de Vakgroep Sociologie`, Katholieke Universiteit Nijmegen, Postbus 9108, 6500 HK Nijmegen, tel. 080-515568

[2]Met dank aan Onno Boonstra, Bert Felling, Peer Meurkens, Vincent Peters, Peer Scheepers,en Theo van der Weegen, voor hun stimulerend commentaar op eerdere versies van dit artikel.

wilt immers zo snel mogelijk naar een analyse met statistische technieken, en daarvoor is een rechthoekige matrix een vereiste.

We willen in dit artikel laten zien, hoe een andere methode van gegevens-opslag, namelijk die volgens het relationele model, een aantal problemen in de dataverzameling en -opslag op kan lossen, zonder dat de mogelijkheden van statistische analyse beperkt worden. Ook zullen we aangeven dat de tijd rijp is om de diverse statistische software aan te passen aan een relationele opslag van data.

Het onderscheid tussen een relationeel en conventioneel model van opslag van gegevens is het thema van het eerste hoofdstuk. We beginnen met een uiteenzetting van het conventionele model. Vervolgens wordt aan de hand van een fictief survey-onderzoek het relationele model toegelicht. Hierbij passeren enkele kernbegrippen en toepassingen de revue. Tenslotte wordt beschreven in hoeverre en in welke zin het relationele model afwijkt van het conventionele. Aldus wordt in algemene zin het nut van een relationeel model aangetoond.

In hoofdstuk 2 illustreren we het gebruik van het relationeel model in een konkreet onderzoek. We hebben gekozen voor een type onderzoek, waarin het werken met verschillende typen data niet vreemd is: de sociale-netwerkanalyse. We zullen zien dat dan de opslag volgens het relationele model een verrijking biedt ten opzichte van bestaande modellen.

1 TWEE METHODES VAN GEGEVENS-OPSLAG

1.1. De conventionele methode

De conventionele manier van gegevens-opslag ten behoeve van sociaal-wetenschappelijk en kwantitatief historisch onderzoek gaat volgens het principe van de rechthoekige datamatrix. Het kenmerkende van zo'n matrix is dat van te voren vastgestelde kenmerken per observatie-eenheid worden gemeten en opgeslagen: "De datamatrix geeft een beschrijving van de onderzoekselementen in termen van de onderzochte eigenschappen op een zodanige wijze, dat er een rechthoekige tabel ontstaat waarbij elke regel een onderzoekselement voorstelt, met daarachter de waarde van de verschillende onderzochte eigenschappen" (Segers,1977,p.94).

Een rondgang door enkele veel gebruikte handboeken over methoden van sociaal-wetenschappelijk en kwantitatief historisch onderzoek leert al snel dat de ingevulde datamatrix als eindfase van de dataverzameling wordt beschouwd, en tevens als beginpunt van de data-analyse (Segers, 1977, p.95; Swanborn,

1981, p.38; Lindblad, 1984, p.26). Of het nu gaat om een survey naar levensbeschouwelijke waarden, om een onderzoek naar de familie-omstandigheden te Gent in 1925, of om de vraag welke factoren van invloed zijn op het weer aan de slag komen van werklozen, in al deze gevallen worden de gegevens geclassificeerd naar waarnemings- of onderzoekseenheid en naar variabele. De resultante hiervan is de rechthoekige datamatrix met als belangrijkste kenmerk de "ondubbelzinnige plaatsbepaling van elk getal in dit stelsel" (Lindblad,1984, p.26). De rechthoekige datamatrix leent zich voor een veelvoud van bewerkingen, van eenvoudige univariate tellingen tot complexe multivariate analyse.

Als we de methodenboeken mogen geloven, valt er niet te ontkomen aan de opslag van gegevens in een rechthoekige datamatrix, niet bij kwantitatief survey-onderzoek, maar ook niet bij kwalitatief sociologisch en historisch onderzoek. De datamatrix is blijkbaar niet alleen beginpunt van statistische analyse, maar tevens begin- en eindpunt van de fasen van dataverzameling en -opslag. Een andere structuur lijkt onmogelijk, wordt althans in de door ons geraadpleegde handboeken niet onderkend, meestal met het argument dat dan statistische analyse niet mogelijk zijn.

1.2. Het relationele model en het conventionele model in de praktijk

Om de essentie te laten zien van een relationeel model gaan we uit van een fictief survey-onderzoek. In dit survey is aan een groot aantal respondenten uit enige tientallen verschillende gemeenten een vragenlijst voorgelegd. De eerste set van vragen heeft betrekking op kenmerken van de respondenten als individuen, zoals leeftijd, beroep, religie, sexe, lidmaatschappen van verenigingen, arbeidspositie en dergelijke. De tweede reeks betreft het oordeel van de respondenten over kwesties op uiteenlopende zaken als moraal, ethiek, politiek, religie, sociaal leven en vrijetijdsbesteding. Doel van zo'n survey is grofweg samenhangen te ontdekken tussen bepaalde categorieën personen en hun opvattingen over de genoemde kwesties: denken ouderen anders over ethische vraagstukken dan jongeren, is religieuze affiliatie, sexe, of geografische herkomst daarop van invloed? Hangen bepaalde visies op de kwaliteit van het sociale leven samen met denkbeelden over economische politiek? In het algemeen vindt men deze verbanden via rechtstreekse bevraging van het individu. Anders wordt het met samenhangen als: zijn mensen woonachtig in een agrarisch milieu andere meningen toegedaan dan stadsbewoners? Welk type vereniging speelt een rol in de vrijetijdsbesteding van mensen? Om vragen als de laatste twee te kunnen beantwoorden is het noodzakelijk omgevingskenmerken (gemeenten en verenigingen) aan individuen te koppe-

len. Alle respondenten uit Nijmegen bijvoorbeeld krijgen de urbanisatiegraad, de religieuze samenstelling en de vigerende politieke verhoudingen in de gemeenteraad als individuele kenmerken toegedacht. Ook inzake verenigingen is dit het geval: classificaties naar doelstelling en reikwijdte worden als kenmerken van individuen beschouwd.
In algemene zin is het doel van de analyses de statistische spreiding of samenhang van kenmerken over en tussen groepen van individuen vast te stellen. De feitelijke waarnemingseenheid is een individu; de variabelen daarentegen hebben betrekking op uiteenlopende sociale eenheden: personen,gemeenten en verenigingen.

De conventionele structuur. Indien we bovenstaande gegevens willen opslaan in één rechthoekige matrix, die we als kenmerkend beschouwen voor de conventionele methode, dan levert dit een aantal nadelen op. Een drietal wordt hieronder besproken onder de noemers keuzemomenten, opslagruimte en interne consistentie.

We komen voor een aantal keuze-momenten te staan:
- Tevoren moet vastgesteld worden van hoeveel verenigingen of van welke verenigingen een respondent maximaal lid kan zijn, omdat voor iedere respondent evenveel gegevens opgeslagen moeten kunnen worden. De keuze van dit maximum is altijd arbitrair.
- Voordat we de gegevens gaan invoeren, moet voor de gemeenten bepaald worden welke gegevens er meegenomen worden (type gemeente, aantal inwoners e.d.) Voor bepaalde gegevens moet vooraf ook een classificatie-schema gemaakt zijn (b.v. urbanisatiegraad). Achteraf aanpassen van dit schema is slechts mogelijk binnen het kader van het schema waarmee we begonnen zijn. Ook het verbeteren van de matrix is dan een gecompliceerde en riskante bezigheid: voor iedere respondent moet bekeken worden in welke gemeente deze woonachtig is, vervolgens moeten dan deze gemeentegegevens op korrekte manier gewijzigd worden.

Veel gegevens komen meerdere malen in het bestand voor. Er wordt daardoor op een minder efficiënte wijze gebruik gemaakt van opslagruimte op het achtergrondgeheugen van de computer.
- Kiezen we bij de variabele 'lidmaatschap van verenigingen' bijvoorbeeld voor een hoog maximum aantal, dan zal veel gereserveerde ruimte voor respondenten met een laag aantal lidmaatschappen onbenut blijven.
- Indien we ook informatie over verenigingen willen opslaan, (doelstelling) dan zal voor elk lidmaatschap van een respondent deze informatie opgeslagen moeten worden. Veel respondenten zullen lid zijn van dezelfde verenigingen,

zodat deze gegevens herhaaldelijk in de datamatrix voorkomen. Voor de gegevens van de gemeenten waarin men woonachtig is geldt hetzelfde.

Met name bij micro-computers zal de omvang van een bestand parten gaan spelen in de fase van data-opslag.

Doordat kombinaties van gegevens meer dan eens moeten worden opgeslagen, loopt de interne consistentie van de data-matrix gevaar. In ons voorbeeld is dit met name het geval bij de gegevens over verenigingen en gemeenten:

- De kombinatie gemeente en urbanisatiegraad komt meerdere malen voor. Daardoor kunnen fouten makkelijk insluipen. Bijvoorbeeld: de ene keer krijgt de gemeente Nijmegen urbanisatiegraad 1 en de andere keer, als gevolg van een type-fout of een verkeerd doorgevoerde aanpassing, graad 7. Hetzelfde geldt voor verenigingen die vaker vermeld worden.

Een structuur van gegevens-opslag waarin kombinaties van gegevens frequent voorkomen, zoals in de conventionele rechthoekige datamatrix, is gevoelig voor dergelijke inconsistenties.

Het relationele model. De relationele structuur van gegevensopslag kent voor bovenstaande problemen een oplossing. Door de gegevens op te slaan volgens de structuur-regels van dit model ontstaat een database die onder meer de volgende voordelen biedt:

- Het is niet noodzakelijk om vooraf te bepalen van hoeveel of welke verenigingen de respondenten lid kunnen zijn.
- Gegevens over organisaties en verstedelijking komen per organisatie en per gemeente slechts eenmaal voor. De interne consistentie van de gegevens wordt door de structuur gewaarborgd.
- Doordat deze gegevens slechts eenmaal opgeslagen worden, wordt bij de opslag van die data veel ruimte bespaard op een computer-geheugen.
- Classificatie-schema's van gemeenten naar urbanisatie-graad of van verenigingen naar sektor kunnen elk moment als individuele kenmerken aan de respondenten worden toegevoegd.

Na de introductie van het relationele model door Codd (Codd,1970), is men in de administratieve informatievoorziening steeds meer overgegaan tot de implementatie van dit model bij de opslag en het onderhoud van een database (Date 1983, 1985). Het relationele model is deels ontleend aan de verzamelingenleer, deels aan de relationele algebra. Hierna passeren enkele begrippen de revue, voorts demonstreren we hoe een relationele database opgezet kan worden.

begrippen

Alle gegevens worden in het relationele model beschouwd als unieke kenmerken van onderzoeksobjekten. Ze worden opgeslagen in wat men 'relaties' noemt (in dit geval een equivalent voor matrix of bestand). Een relatie bestaat uit kolommen(kenmerken of variabelen) en rijen (de verschijningsvorm van een reeks kenmerken in de werkelijkheid; het onderzoeksobject: individuen, gemeenten en verenigingen). In relaties worden kenmerken met elkaar in verband gebracht op een zodanige wijze dat iedere rij een unieke verschijningsvorm heeft. Bovendien mag die verbinding van kenmerken (persoon en gemeente) geen als boven omschreven inconsistenties opleveren.

De opslag van gegevens geschiedt in een relationele structuur ook in een rechthoekige matrix. In de regel is er echter sprake van meerdere matrices (relaties); in elke matrix worden bepaalde kenmerken vermeld. Deze zijn onlosmakelijk met elkaar verbonden en verwijzen per definitie naar elkaar. (Voorbeeld: een relatie met gemeente-gegevens bestaat uit de kenmerken gemeentecode,gemeentenaam, urbanisatiegraad; de beide laatste horen per definitie bij de gemeentecode en iedere gemeentecode krijgt een naam en een urbanisatiegraad toebedeeld)

Naast relatie is sleutel een centraal begrip. Een sleutel is een kenmerk met twee uiteenlopende functies :

- Het zorgt ervoor dat iedere rij in een relatie als uniek te identificeren is.
- Het maakt een verbinding mogelijk met andere relaties, zodat gegevens uit verschillende matrices met elkaar gekoppeld kunnen worden.

Essentieel in een relationele structuur is dat ieder kenmerk in een relatie "een feit moet zijn over de sleutel, de gehele sleutel en niets anders dan de sleutel" (Date 1985, p. 234). In de structuur dienen geen ongewenste afhankelijkheden voor te komen: indien een invulling van een kenmerk per definitie voorkomt in kombinatie met een ander niet-sleutel-kenmerk, dan hebben we een ongewenste structuur (bijvoorbeeld: indien we de gemeente van de respondent weten dan heeft de urbanisatiegraad per definitie een bepaalde waarde). De interne consistentie van de gegevens wordt dan niet meer door de structuur gegarandeerd. In praktijk wordt dit opgelost door de structuur te 'normaliseren'. Alleen daarna garandeert de structuur een interne consistentie, voorkomt men het veelvuldig opslaan van dezelfde gegevens (dit noemt men ook wel het vermijden van redundantie), en houdt men de structuur open voor eenvoudige en eenduidige wijzigingen.

structuur: normaliseren

Grofweg valt normaliseren te omschrijven als: het terugbrengen van een werkelijkheid van komplexe verbanden tot een aantal relaties waarbinnen slechts

één verband bestaat: dat van een kenmerk tot de sleutel van de relatie. We zullen nu stapsgewijs aangeven hoe normaliseringsprocedures worden toegepast in ons survey-onderzoek, nog voordat de gegevens in tabellen worden opgeslagen.

Het normaliseren begint met het op één rij zetten van alle gegevens die voor een onderzoeksvraag van belang zijn. Alle kenmerken worden als één grote verzameling gedefinieerd. Een dergelijke verzameling van kenmerken wordt als volgt aangegeven:

respondent(resp#,leeftijd,gemeente,vraag1...vraagn,vereniging,sektor,organisatiegraad,
urbanisatiegraad,gemeentepolitiek, gemeentereligie)

Zouden we alles vastleggen volgens deze structuur, dan zijn we bezig volgens de conventionele methode: alles in één rechthoekige matrix. Het relationele model vraagt om een verdere aanpassing en uiteenlegging. Allereerst moet er op gelet worden dat alle te onderzoeken kenmerken vermeld zijn, dat een kenmerk niet meerdere malen in de rij voorkomt, en dat kenmerken uitgesplitst worden tot de kleinst mogelijke nog relevante eenheid. Ook moet een sleutel worden opgenomen in bovengenoemde verzameling die elke rij als uniek kan identificeren: veelal gebruikt men daarvoor het respondentnummer.

Is een verzameling kenmerken op bovenstaande manier geordend, dan kan de structuur aan de eisen van een relationeel model worden aangepast. Dit aanpassen verloopt in stappen. Elke stap brengt de structuur dichter bij het ideale model. We beginnen met de zogenaamde eerste normaalvorm.

eerste normaalvorm

Een relatie staat in de eerste normaalvorm indien alle kenmerken slechts enkelvoudig voorkomen. Kunnen kenmerken meervoudig voorkomen (bijvoorbeeld lidmaatschappen van verenigingen), dan worden deze afgesplitst en in een aparte relatie geplaatst, uiteraard inclusief de sleutel. Het verband tussen beide relaties (matrices) kan dan korrekt gelegd worden. Als we bijvoorbeeld de verenigings-gegevens afsplitsen van de oorspronkelijke gegevens, komt de volgende structuur tot stand:

respondent(resp#,leeftijd,gemeente,vraag1...vraagn, urbanisatiegraad,gemeentepolitiek, gemeentereligie)

lidmaatschap(resp#,vereniging,sektor,organisatiegraad)

De tweede relatie bevat slechts kenmerken die betrekking hebben op het lidmaatschap van een vereniging. In deze nieuwe relatie is de sleutel een kombinatie van respondentnummer en vereniging. Een van beide is niet voldoende.

Door deze afsplitsing van kenmerken is al aan twee nadelen van de enkelvoudige matrix tegemoet gekomen:

- Het aantal lidmaatschappen hoeft niet tevoren vastgelegd te worden. Een respondent mag van een onbeperkt aantal verenigingen lid zijn.
- Een efficiëntere opslag van gegevens is mogelijk omdat er minder lege ruimte ontstaat in het bestand.

Indien er geen meervoudige kenmerken meer aanwezig zijn in de structuur, dan staan de relaties in de eerste normaalvorm. We kunnen dan overgaan tot de tweede normaalvorm:

tweede normaalvorm

Een relatie staat in de tweede normaalvorm indien alle kenmerken volledig afhankelijk zijn van de gehele sleutel. Kenmerken die slechts van een deel van de sleutel afhankelijk zijn worden afgesplitst, inclusief het daarbij horend deel van de sleutel. In deze stap zijn dus slechts relaties in het geding met een sleutel die uit meerdere kenmerken bestaat.

In het voorbeeld moeten we dus nu alleen de relatie lidmaatschap bekijken. We zien hier dat er sprake is van bovenbedoelde gedeeltelijke afhankelijkheid: sektor en organisatiegraad zijn wel afhankelijk van vereniging maar niet van resp#. We splitsen als volgt:

respondent(resp#,leeftijd,gemeente,vraag1...vraagn,urbanisatiegraad,gemeentepolitiek, gemeentereligie)

lidmaatschap(resp#,vereniging)

vereniging(vereniging,sektor,organisatiegraad)

Door deze afsplitsing zijn drie nadelen van de eerste normaalvorm verdwenen.

- Er is geen inconsistentie mogelijk tussen vereniging en sektor. Elke vereniging wordt slechts eenmaal opgeslagen en daaraan wordt slechts eenmaal een sektor gekoppeld.
- Het slechts eenmaal opslaan van elke vereniging levert wederom ruimte-winst op.
- De sektor-indeling en organisatiegraad kunnen zo nodig veranderd worden. We hoeven dus niet persé vooraf te kiezen voor een bepaalde indeling.

Met deze afsplitsing staan onze gegevens in de tweede normaalvorm. We mogen nu overgaan tot de derde normaalvorm:

derde normaalvorm

Een relatie staat in de derde normaalvorm indien alle kenmerken alleen van de sleutel afhankelijk zijn, en niet van andere gegevens binnen de relatie.

In ons voorbeeld is er sprake van een dergelijke afhankelijkheid binnen de respondenten-relatie: urbanisatiegraad, gemeentepolitiek en gemeentereligie zijn afhankelijk van zowel resp# als van (en eigenlijk nog meer van) gemeente. Afsplitsing van deze gegevens levert de volgende structuur op:

respondent(resp#,leeftijd,gemeente,vraag1...vraagn)

gemeente(gemeente,urbanisatiegraad,gemeentepolitiek, gemeentereligie)

lidmaatschap(resp#,vereniging)

vereniging(vereniging,sektor,organisatiegraad)

Met een stippellijn wordt aangegeven dat het kenmerk gemeente in de relatie respondent elders een sleutel-gegeven is. De stap naar de derde normaalvorm is in feite een aanscherping van de eisen die aan de eerste en tweede normaalvorm gesteld worden:

- Gemeente-kenmerken worden nu slechts eenmaal per gemeente opgeslagen, zodat de interne consistentie vanuit de structuur gegarandeerd wordt.
- Gemeente-kenmerken zijn moeiteloos tussentijds aan te passen. Een classificatie van gemeenten kan zelfs achteraf aangebracht worden.
- Door het slechts eenmaal opslaan van deze gegevens besparen we veel ruimte (zeker bij grote aantallen gemeenten en respondenten).

De gegevens staan nu in de derde normaalvorm, en de eerder genoemde bezwaren gelden nu niet meer. Verdere normalisatie is daarom niet meer nodig. In de praktijk werkt dit ook vaak zo: na een aantal normalisatie-stappen is de structuur geoptimaliseerd en kan de opslag van gegevens beginnen.

Tot zover het principe van een relationeel model. Het resultaat is een structuur die recht doet aan het feit dat bij een onderzoek kenmerken of variabelen op meerdere typen onderzoeks-objekten betrekking hebben. De werkelijkheid wordt adequater en flexibeler weergegeven door gegevens van ieder type onderzoeksobject afzonderlijk op te slaan. Samenvattend levert dat in ons voorbeeld de volgende voordelen op:

- De opslag van redundante gegevens wordt tot een minimum beperkt.
- De structuur garandeert een interne consistentie van de gegevens.
- Keuzes met betrekking tot het opslaan van kenmerken van bepaalde deelverzamelingen kunnen tot op het laatste moment uitgesteld worden.

We denken hiermee voldoende aangetoond te hebben dat opslag van data volgens het relationele model ook voor sociaal- wetenschappelijk onderzoek van nut kan zijn.

opvragen van gegevens: vraagtaal

De structuur die met normalisatie wordt beoogd, is afhankelijk van de onderzoeksvragen. De uitsplitsing van lidmaatschappen is bijvoorbeeld nodig als we willen weten van welk type vereniging respondent lid is.

Met behulp van een zogenaamde vraagtaal kunnen de opgeslagen gegevens in allerlei combinaties en nuances worden opgevraagd. Het gaat hierbij meestal om een aantal operaties die zijn ontleend aan de verzamelingenleer (unie, vereniging, intersectie, verschil en produkt) en de relationele algebra (projectie, selectie, join). In het vervolg van dit artikel maken we bij de voorbeelden gebruik van een algemene syntax-beschrijving van de taal SQL (Structured Query Language). Deze taal staat dicht bij operaties zoals ze in de verzamelingenleer zijn gedefinieerd en is voor iedereen zonder veel uitleg te lezen.[3] Voor rekenwerk leent deze taal zich nauwelijks.

Een veel voorkomende procedure is het in een rij plaatsen van alle noodzakelijke gegevens per respondent. Deze rechthoekige datamatrix kan vervolgens met statistische software-pakketten geanalyseerd worden. Een dergelijke samenvoegingsprocedure ziet er in ons voorbeeld als volgt uit (de relaties zijn in het voorbeeld voor het gemak afgekort tot de eerste letter):

```
SELECT resp#,leeftijd,gemeente,vraag1...vraagn,vereniging,sektor,organisatiegraad,urbanisatiegraad,
       gemeentepolitiek,gemeentereligie
FROM   Respondenten,Lidmaatschappen,Verenigingen,Gemeenten
WHERE  R.resp#=L.resp#
AND    L.vereniging=V.vereniging
AND    R.gemeente=G.gemeente
```

In deze omvangrijke rechthoekige datamatrix zijn de 4 relaties (matrices) uit ons voorbeeld samengevoegd. De sleutels zorgen ervoor dat de juiste kenmerken bij de juiste respondenten terecht komen. Door meer voorwaarden te stellen kan men alvast een data-selectie toepassen op de respondentgegevens (b.v. AND urbanisatiegraad =1 levert een matrix op met respondenten uit gemeenten met urbanisatiegraad 1).

Het zou voor sociale wetenschappers een stuk gemakkelijker worden, indien statistische pakketten al rekening houden met een relationele opslag. Een tussenstap via SQL is dan niet meer noodzakelijk. Sommige pakketten, zoals SAS, bieden al deze mogelijkheid, zij het nog niet op een gebruiksvriendelijke manier.

[3] Zie voor meer informatie: (Date 1983, 1985, Janssens 1984, van der Lans/van Stegeren 1985, Reitsma 1985, Langendoen 1985, Jacobs/Vanparys 1986)

1.3. Het relationele model: een alternatief

We hebben laten zien dat in een doorsnee-onderzoek het relationele model voor gegevensopslag voordelen biedt. We krijgen een adequatere opslag voor bepaalde vraagstellingen, bijvoorbeeld met betrekking tot omgevingskenmerken (urbanisatiegraad, religieuze omgeving) of sociale leven (verenigingen).

Relationele databases zijn wellicht minder interessant voor surveys dan voor onderzoek waarbij per definitie meerdere typen onderzoeksobjekten in het geding zijn (bijvoorbeeld individuen, gezinnen, gemeenten en gebeurtenissen). Bij veel onderzoekingen komt echter al gauw meer kijken dan een eenvoudige vragenlijst met per vraag één antwoord. Vooral wanneer de relaties tussen verschillende soorten objekten eveneens onderwerp van onderzoek zijn, biedt een relationeel model veel mogelijkheden. In bepaalde typen onderzoek wordt al gewerkt met een relationele structuur. Het is dan vanzelfsprekend om uit te gaan van meerdere matrices die aan elkaar gekoppeld worden. Voorbeelden zijn kwalitatieve analyses op meerdere nivo's, contextuele analyse en sociale-netwerkanalyse. In veel onderzoekingen wordt echter (nog) afgezien van de relationele opslag, omdat men in de veronderstelling verkeert dat statistische analyses enkel via opslag in een enkelvoudige matrix mogelijk zijn. Het bovenstaande voorbeeld heeft naar onze mening deze veronderstelling voldoende ontkracht.

In het volgende deel zullen we zien dat ook in analyse-modellen, waarin al rekening wordt gehouden met de koppeling van meerdere matrices, het relationele model een verrijking aan mogelijke bewerkingen biedt.

2. EEN ONDERZOEKSVOORBEELD: DE KATHOLIEKE BEWEGING IN ARNHEM

In dit hoofdstuk worden enkele aspekten van een historisch-sociologisch onderzoek besproken, met als specifieke invalshoek het gebruik van en het werken met relationele databases ten behoeve van sociaal-netwerkanalyses.

2.1. Probleemstelling

Het verloop van de katholieke beweging te Arnhem vormt onderwerp van onderzoek. Het onderzoek is erop gericht inzicht te krijgen in de dynamiek van het verloop van de beweging op de lange termijn. Daarbij wordt gelet op drie onderling samenhangende processen: organisatorische ontwikkelingen, mobiliserings-processen en ideologie (Schreuder, 1981). Organisatorische processen als schaalvergroting en specialisatie worden onderzocht op hun effecten voor de interne communicatie. Het netwerk van bestuurlijke betrekkingen tussen organisaties door middel van dubbelfuncties van bestuurders is voorwerp van analyse in dit Arnhemse onderzoek. Mobiliseringsprocessen verlopen meestal

via al bestaande, hechte relatienetwerken (familie en vrienden). Het idee is verder dat het draagvlak van de beweging wordt versterkt en de kans op een succesvolle mobilisering wordt groter, naarmate tussen de diverse geledingen van de beweging (personen en organisaties) hechte netwerken van relaties bestaan en worden gekreëerd.

Ook in het licht van mobiliseringsprocessen zijn sociale netwerken dus relevante onderzoeksobjecten. Daarom wordt in dit onderzoek de aandacht gevestigd op de mate en betekenis van de overlap van bestuurlijke en familiale netwerken en op de differentiële samenstelling van organisationele en bestuurlijke netwerken.

Alvorens te starten met een bespreking van de verzameling, opslag en analyse van de Arnhemse gegevens, eerst enkele algemene opmerkingen over sociaal-netwerkanalyse als onderzoeksinstrument.

2.2. Sociaal-netwerkanalyse; begrippen,thema's en toepassingen

Het begrip sociaal netwerk refereert in een alledaagse betekenis aan de relaties die een persoon, organisatie, bedrijf, of politieke partij heeft, en die hij of zij kan aanwenden om een bepaald doel te bereiken of boodschap over te brengen. In abstracto is een netwerk een structuur van relaties tussen bepaalde eenheden. Netwerkanalyse is een op de wiskundige graphentheorie gebaseerd hulpmiddel om onder meer deze structuur bloot te leggen, en de positie van de afzonderlijke eenheden te bepalen (Felling, 1974).

Het gebruik van het netwerkconcept veronderstelt gegevens die gearrangeerd kunnen worden als punten die door lijnen met elkaar verbonden worden (Plakans, 1984, p.132). Het is aan de onderzoeker om aan te geven waar de punten en lijnen in zijn onderzoek voor staan. Een punt in een netwerk kan staan voor een individu, groep, bedrijf of familie, de lijn kan staan voor een familiebetrekking, de uitwisseling van informatie, het verstrekken van goederen en financiën, of het vervullen van een functie.

Het moge duidelijk zijn dat het gebruik van netwerkanalyse eisen stelt aan de opslag van gegevens. Er moet sprake zijn van minstens twee verzamelingen (punten en lijnen), die onderling gekoppeld kunnen worden. De verzameling punten bestaat bijvoorbeeld uit de leden van een familie, terwijl de lijnen informatie geven over de specifieke verwantschapsrelatie tussen elk tweetal familieleden. Aldus ontstaan netwerken van individuen: unipartite netwerken. Een ander type netwerk is het bipartite netwerk. Het kenmerkende hiervan is dat er twee verschillende puntenverzamelingen en een lijnenverzameling in het geding zijn, bijvoorbeeld personen en organisaties. De lijnen in een dergelijk netwerk geven aan in welke organisatie een persoon functies vervuld heeft.

2.3. Verzameling en opslag van de gegevens

Over twee soorten sociale eenheden zijn in het Arnhemse onderzoek gegevens verzameld: organisaties en personen. Er zijn grofweg drie mogelijkheden voor de opslag van deze gegevens. De eerste mogelijkheid is één grote matrix met daarin alle gegevens. Teneinde recht te doen aan het longitudinale karakter van de studie zou deze matrix voor een aantal vooraf geselecteerd peiljaren aangemaakt moeten worden. Stel dat de keuze valt op vier peiljaren dan betekent dit vier bestanden met daarin gegevens over personen, organisaties en functies. Het spreekt voor zich dat het gescheiden opslaan van gegevens over deze drie onderzoekseenheden de voorkeur heeft. De tweede mogelijkheid is het werken met 12 bestanden: voor elk peiljaar één voor personen, één voor organisaties en één voor functies. Het op netwerkanalyse afgestemde programma GRADAP nodigt uit tot deze opzet (Stokman, Van Veen,1981). Het derde alternatief is opslag volgens het relationele model. Dit leidt tot drie aparte bestanden: personen, organisaties en functies.

De keuze is in feite beperkt tot de twee laatste mogelijkheden. In beide gevallen is het aantal door één persoon te vervullen functies in principe onbeperkt; ook het bestuur kan uit een onbeperkt aantal personen bestaan. De voordelen van het relationele model reiken echter verder. Een voordeel is dat vooraf geen beslissingen genomen hoeven te worden over de keuze van peiljaren, de lengte van de records, de codering van de kenmerken en dergelijke. De flexibiliteit van het relationele databestand is groot. Door het opnemen van tijdsvariabelen, zoals bijvoorbeeld oprichtings- en opheffingsjaar van de organisaties, geboortejaren van bestuurders en begin- en eindjaar van functies, is het mogelijk andere jaren in de analyse te betrekken, of om het bestand aan te vullen met gegevens over recente ontwikkelingen. Toevoeging van deze kenmerken maakt de structuur van een relatie universeel in tijd: men hoeft geen rekening meer te houden met de jaren waarover men berekeningen en metingen wil uitvoeren. Een tweede voordeel is dat er direkte computergestuurde controles mogelijk zijn op de invoer, tevens kan de interne consistentie van de verzamelde gegevens bewaakt worden.

De relationele structuur is bovendien beter afgestemd op de onderzoekspraktijk. Gegevens over netwerken worden vaak uit diverse bronnen en archieven geput, en op uiteenlopende tijdstippen verzameld. Soms worden gegevens verzameld via het domein van de organisatie (bijvoorbeeld jaarverslagen), soms via de persoon (curriculum vitae, Who is Who's). De onderzoeker kan beide soorten gegevens zonder noemenswaardige voorbewerking opnemen in zijn database. Hij begint met de functiegegevens in te voeren, en door ingebouwde controles wordt hij automatisch geïnformeerd of de betreffende personen of organisaties al "bestaan", of er niet al identieke identificatienum-

mers gekozen zijn, of een persoon een functie uitoefent terwijl hij reeds als overleden te boek staat, en wat dies meer zij. Bovendien biedt het mogelijkheden voor longitudinaal onderzoek.

Hoe zijn nu in het Arnhemse onderzoek de gegevens over personen en organisaties verzameld en opgeslagen?

Organisaties. Alle in Arnhem voorkomende katholieke organisaties zijn geïnventariseerd. Vervolgens zijn allerlei kenmerken van deze organisaties verzameld, zoals jaar van oprichting en opheffing, juridische structuur, sector etc. Al deze informatie is opgeslagen in een bestand ORGANISATIES (Onummer, Onaam, Oprichting, Opheffing, Sektor, overige kenmerken).

Personen. In eerste instantie zijn de namen van de bestuurders van bovenbedoelde organisaties verzameld, inclusief informatie over de aard en de duur van de functie. Daarna is een keur aan biografische gegevens verzameld over ongeveer 700 personen. Op basis hiervan wordt de katholieke voorhoede in Arnhem beschreven. Deze biografische gegevens zijn ondergebracht in het bestand PERSONEN (Pnummer, Pnaam, geslacht,kerkelijke staat, geboortejaar en -plaats, etc) De diverse functies zijn ondergebracht in het bestand FUNCTIES (Pnummer, Onummer, Beginjaar, Eindjaar, Functieniveau)

Van alle personen zijn de onderlinge familierelaties vastgesteld en in een bestand FAMILIE opgeslagen. Alle punten in dit netwerk zijn personen (Pnummer, Pnummer2, soort relatie).

In het bestand FINANCIEN is van elke persoon zo nodig informatie opgeslagen over zijn financiële betrokkenheid bij de Arnhemse katholieke beweging. Dit bestand is qua structuur vergelijkbaar met FUNCTIES, maar uiteraard gevuld met andersoortige informatie (Pnummer, Onummer, Aard financiële betrokkenheid, Jaar c.q.periode van gift (jaargift), Bedrag).

Samenvattend: ten behoeve van het Arnhemse onderzoek zijn 6 verschillende databestanden opgebouwd. Onderstaand overzicht geeft nogmaals de namen en de sleutel waarmee ze onderling gekoppeld kunnen worden.

ORGANISATIES	PERSONEN	FUNCTIES	FAMILIE	FINANCIEN
Onummer	Pnummer	Pnummer	Pnummer	Pnummer
		Onummer	Pnummer2	Onummer

Door het opnemen van diverse tijdsvariabelen (oprichting en opheffing van een organisatie, geboortejaar van personen, functieduur, en jaar van schenking) is met deze bestanden een longitudinale analyse mogelijk.

Er is niet gekozen voor een opslag van alle relevante gegevens in één tabel, maar meteen gewerkt met een structuur van onderling via sleutels

(unieke identificatie van personen en organisaties) te verbinden relaties. Een fase van normaliseren is voorafgegaan aan de feitelijke opslag.
Namen, beroepen en sektoren zijn slechts eenmaal per persoon of organisatie opgeslagen, en worden in het functiebestand dan ook niet genoemd. Met behulp van de sleutels Pnummer en Onummer is een allesomvattend bestand wel te maken.

2.4. Analyse

In eerste instantie zijn er per peiljaar overzichten gemaakt van personen en hun functies, met vermelding van hun biografische gegevens. De bestanden PERSONEN en FUNCTIES zijn via Pnummer gekoppeld, en via gegevens over de functieduur zijn uiteindelijk de peiljaarselecties gemaakt. Mochten deze peiljaren geen duidelijk antwoord geven op het gestelde probleem, dan gaat het voordeel van de relationele structuur pas goed gelden: andere peiljaren kunnen immers gekozen worden.

In termen van de verzamelingenleer gaat het hier om een vereniging van drie verzamelingen tot een nieuwe onder bepaalde voorwaarden. De eerste stap betreft dan het vastleggen van de functies uit een bepaald peiljaar. In de vraagtaal ziet dat er als volgt uit.

```
SELECT Pnummer,Onummer,Fgegevens
FROM   functies
WHERE  (beginjaar<=peiljaar)
  AND  ((eindjaar>=peiljaar) OR (eindjaar='niet ingevuld'))
```

Voor ieder peiljaar levert dit een selectie op van alle bestuurlijke functies. We noemen deze selectie in het vervolg de PEILJAARSELECTIE. Na deze selectie is als tweede stap informatie uit andere bestanden te verenigen tot een groot bestand voor het maken van statistische overzichten:

```
SELECT Pnummer,Onummer,Ogegevens,Pgegevens,Fgegevens
FROM   personen,organisaties,functies
WHERE  Pnummer IN (PEILJAARSELECTIE)
  AND  Onummer IN (PEILJAARSELECTIE)
```

Een dergelijke vereniging had ook in één keer gekund, zonder de voorafgaande peiljaarselectie. Omdat deze selectie in het vervolg vaker gebruikt zal worden is toch gekozen voor een voorselectie op peiljaar. In het Arnhemse onderzoek zijn de resultaten hiervan opgeslagen in aparte hulpbestanden, omdat de aanmaak van een dergelijke selectie voor iedere operatie te veel tijd zou kosten.

De biografische gegevens zijn vervolgens per peiljaar in enkele frequentietabellen samengevat. Ook is nagegaan of de per peiljaar geselecteerde personen onderling verwant zijn, en of zij in de periode rond het betreffende peiljaar de beweging op een of andere wijze financieel hebben ondersteund.

Verwantschap en financiële giften per peiljaar zijn weer verenigingen onder voorwaarde: verenig de PEILJAARSELECTIE met de bestanden FAMILIE en FINANCIEN onder voorwaarde dat het persoonsnummer voorkomt in beide verzamelingen. Bij FAMILIE hoort daar nog als extra voorwaarde bij dat ook de tweede persoon moet voorkomen in de PEILJAARSELECTIE:

```
SELECT Pnummer,Pnummer2
FROM   familie
WHERE  Pnummer  IN (PEILJAARSELECTIE)
  AND  Pnummer2 IN (PEILJAARSELECTIE)
```

Voor FINANCIEN geldt als extra voorwaarde dat de gift in een periode rondom het peiljaar moet liggen:

```
SELECT Pnummer,gift
FROM   financien
WHERE  Pnummer IN (PEILJAARSELECTIE)
  AND  ((jaargift<=peiljaar+5)
  AND   (jaargift>=peiljaar-4)
```

Met behulp van al deze onderzoeksgegevens zijn, om het in de traditionele onderzoekstaal te stellen, de kaartenbakken op een zodanige wijze door elkaar geschud dat er per peiljaar voldoende systematische gegevens beschikbaar zijn om de eerste onderzoeksresultaten te beschrijven. Over netwerken is dan nog niet eens gesproken.

Een beschrijving van de samenstelling en omvang van de bestuurlijke en organisatorische netwerken per peiljaar is het volgende doel van de analyse. Om te laten zien wat een relationele structuur in combinatie met een eenvoudige vraagtaal in de sfeer van sociaal-netwerkanalyse mogelijk maakt, volgt hieronder een beschrijving van de zogenaamde inductieprocedure. Met behulp van inductie worden bijvoorbeeld de netwerken van respectievelijk bestuurders en organisaties gegenereerd per peiljaar, en hierbinnen nog relevant geachte deelnetwerken van bepaalde categorieën personen en of organisaties.

De inductie-operatie wordt uitgevoerd op de matrix die verbanden legt tussen onderzoekseenheden, in ons geval FUNCTIES. Het gaat bijvoorbeeld om de relaties tussen bestuurders als gevolg van functies in een en dezelfde organisatie; de organisatie induceert nu een lijn tussen de bestuurders. Hoe een dergelijke operatie in concreto voor een peiljaar wordt uitgevoerd, blijkt uit de volgende bewerking van het bestuurdersnetwerk. In de geïnduceerde graph zijn de punten personen en de lijnen organisaties. Ieder verband tussen

een bestuurder en een organisatie wordt vergeleken met alle andere verbanden tussen dergelijke punten. In feite is dit een vergelijking van de verzameling functies met zichzelf. Daarom wordt het functies-bestand tweemaal genoemd.

```
SELECT P1.Pnummer,P2.Pnummer
FROM   functies:P1,functies:P2
WHERE  (P1.Onummer=P2.Onummer)
  AND  (P1.Pnummer<>P2.Pnummer)
  AND  P1.Pnummer IN (PEILJAARSELECTIE)
  AND  P2.Pnummer IN (PEILJAARSELECTIE)
```

Van deze (deel)verzamelingen zijn via algemene statistische software ondermeer de volgende kenmerken vastgesteld:

- aantal punten;
- totaal aantal kontakten tussen die punten;
- de multipliciteit van de relaties;
- de inductiekracht per punt, bijvoorbeeld het aantal relaties tussen personen dat door een organisatie gelegd wordt;
- het aantal indirekte kontakten per punt.

Het is in principe mogelijk om allerlei deelnetwerken en partiële netwerken apart te bekijken, bijvoorbeeld alle relaties tussen geestelijken en leken, of alle relaties tussen organisaties die geïnduceerd worden door politici. Ook kan worden nagegaan of er in de diverse peiljaren sprake is geweest van een overlap van het bestuurlijke netwerk van personen met hun onderlinge verwantschapsnetwerk. Gegevens hierover zijn te verkrijgen door vereniging van de lijnen die het gevolg zijn van de inductie-selectie met de basisgegevens in PERSONEN en FAMILIE. Het FAMILIE-bestand is immers ook een bestand van lijnen tussen personen maar dan via verwantschap.

Dit voorbeeld van de Arnhemse beweging geeft aan welke mogelijkheden een relationele structuur en een daarop afgestemde vraagtaal bieden. Zonder gebruik te maken van gespecialiseerde netwerkprogrammatuur als GRADAP, zijn onderzoeksvragen met betrekking tot de Arnhemse netwerken te beantwoorden. Dat is ook een verdienste van een relationele database-structuur en de daarop afgestemde vraagtaal. Belangrijker is echter dat de opslag van gegevens plaatsvindt op basis van een flexibele, ruimtebesparende en bij de dataverzameling goed aansluitende relationele structuur. Deze flexibiliteit is juist ook een van dé handelsmerken van netwerkanalyse. Plakans drukt dit als volgt treffend uit: "Of the various grouping concepts we have analyzed...., it is the only one that does not require the preselection of the relationships to be studied, while at the same time it permits all relationships to be examined in the course of the inquiry later" (1984, p.240). Een relationele

structuur van data-opslag biedt de optimale voorwaarden voor het realiseren van deze potentiële mogelijkheden van netwerkanalyse.

3 BESLUIT

In zekere zin mogen we sociaal-wetenschappelijke onderzoekers en historici een bepaalde eenkennigheid verwijten. Ze zweren bij een arsenaal aan meetprocedures en analysetechnieken, en maken de dataverzameling en data-opslag daaraan volledig ondergeschikt. Een blik op het erf van de economische wetenschap of op dat van de administratieve systemen had hen al eerder duidelijk kunnen maken dat de opslag van gegevens vooral preciezer en flexibeler georganiseerd kan worden met een relationele structuur. Dat de aandacht voor relationele database-systemen momenteel toeneemt, blijkt uit diverse feiten. Door nationale beleidscommissies op het terrein van de sociaal-wetenschappelijke informatica wordt de implementatie van relationele structuren van de hoogste prioriteit geacht (Stokman, 1985). Statistische software moet relationele databases als input kunnen hebben. Een belangrijke stap in de goede richting zou zijn, indien voor selectie en inlezen van gegevens een statistisch pakket een relationele vraagtaal bevat. Ontwerpers van bepaalde software hebben deze wending al aangekondigd en soms reeds ten dele toegepast.

Gezien de mogelijkheden van relationele databases is het geenszins aan te raden te streven naar diverse specialistische statistische softwarepakketten, die elk op slechts één type datastructuur en onderzoeksmethode zijn afgestemd. Er zijn diverse specialistische programma's beschikbaar, die - al dan niet met vooropgezette bedoelingen - tot op zekere hoogte gebruik maken van een relationeel model. Deze programma's zijn in de regel geschreven voor een specifieke onderzoeksmethode en -toepassing. Een voorbeeld hiervan is het reeds genoemde GRADAP. Dit pakket is ontwikkeld ten behoeve van netwerkanalyse. Andere voorbeelden zijn het kwalitatieve-analyse pakket KWALITAN (Peters,1986) en programma's met betrekking tot inhoudsanalyse (De Graauw, Op de Weegh, Hutjes, 1985,1986/1987). Voor allerlei uiteenlopende toepassingen en methoden zijn op vergelijkbare wijze toegespitste pakketten te ontwerpen met steeds weer andere eisen aan de datastructuur. Naar onze mening verdient het echter de voorkeur te streven naar een algemene en flexibele databasestructuur, met duidelijke gebruiksregels en naar hierop afgestemde algemene statistische software.

In dit artikel hebben we met het Arnhemse voorbeeld laten zien dat met behulp van een relationele database netwerkgegevens op een identieke wijze gestructureerd kunnen worden als met GRADAP. Dit betekent dus dat voor een klassieke sociaal-netwerkanalyse de definitie en selectie van netwerken even-

goed volgens een algemeen relationeel model kan geschieden; voor analyse is uitsluitend GRADAP geschikt. Een relationele database kan in dit verband beschouwd worden als de volgende stap op de weg waarvan GRADAP de eerste stap was.

Afsluitend kunnen we stellen dat in een relationele database op een eenvoudige wijze allerlei verbanden gelegd worden tussen deelverzamelingen. Ieder aspect van de werkelijkheid, dat intrinsiek aanwezig is in de structuur van de relaties (bestanden), is als een afzonderlijk gegeven op te roepen. Bovendien bestaat de mogelijkheid om structuren te creëren die nauwelijks beperkingen opleggen in de dataverzamelings-fase. De structuur kan worden aangepast aan de wijze waarop gegevens verzameld moeten worden, en aan de aard van de bronnen die de onderzoeker ter beschikking heeft. Daarna zijn allerlei manipulaties mogelijk. Dit alles maakt een relationele database tot een krachtig instrument bij verschillende typen onderzoek binnen de alpha- en gamma-wetenschappen. Of en in welke mate een onderzoeker baat heeft bij deze structuur hangt uiteraard samen met doel en aard van zijn onderzoek.

LITERATUURLIJST

L. Breure en T. Verbaan, Database-systemen en historisch onderzoek. Paper ten behoeve van het historici-congres te Utrecht, mei 1986.

E.F. Codd, A relational model of data for large shared databanks. Communications of the ACM 1976 nr.6

C.J. Date, An introduction to database-systems. Reading, Massachusets 1983.

C.J. Date, Database, Een inleiding. Den Haag 1985

T. Duffhues, Een stadsparochie in beweging. De St. Jansparochie te Arnhem, 1895 - 1985. Arnhem 1985.

T. Duffhues, 'Verzuiling en netwerken. Een theoretisch-methodologische verkenning van verzuilingsprocessen op lokaal niveau', in: It Beaken 48(1986), p.234 - 242.

T. Duffhues, A. Felling, J.Roes, Bewegende patronen. Een analyse van het landelijk netwerk van katholieke organisaties en bestuurders 1945-1980. Baarn-Nijmegen 1985.

A.J.A. Felling, Sociaal-netwerkanalyse. Alphen aan den Rijn 1974

C. de Graauw, J. Hutjes, Gebruik van de micro-computer in de sociale wetenschappen: de beoordeling van database-programmatuur voor de invoer van onderzoeksgegevens. MISO-reeks moduul 1, ITS Nijmegen 1985.

C. de Graauw, J. Hutjes, P. op de Weegh: dBASE III in onderzoek. MISO-reeks moduul 2, ITS Nijmegen 1986.

C. de Graauw, J. Hutjes, P. op de Weegh: Framework II in onderzoek. MISO-reeks moduul 3, ITS Nijmegen 1987 (prepublikatie)

B. Jacobs, R. Vanparys, Normaliseren, nog eens aan de orde. Informatie 1986 nr.9 blz. 741-747.

H. Janssens: Principes van relationele databanken I+II. Informatie 1984 nr.7+8 blz. 523-529 607-616.

J.W.M. Langendoen, Normalisatie, een pad met valkuilen. Informatie 1985 nr.3 blz. 194-196.

J. Th. Lindblad, Statistiek voor historici. Muiderberg 1984

A. Plakans, Kinship in the past. An Anthropology of european family life, 1500-1900. Oxford 1984

V. Peters, KWALITAN, een ondersteunend programma bij de analyse van kwalitatieve gegevens. Nijmegen 1987 (voorlopige uitgave)

E.H. Reitsma, Grenzen en beperkingen van SQL. Informatie 1985 nr.3 blz. 198-201.

O.Schreuder, Sociale bewegingen. Een systematische inleiding. Deventer 1981.

J.H.G. Segers, Sociologische onderzoeksmethoden. Assen, Amsterdam 1977.

F.N. Stokman en F.J.M.A. van Veen (ed.), Gradap. Graph definition and analysis package. User"s manual. Amsterdam 1981. 2 dln.

F.N. Stokman en F.J. van Veen, Programmatuurontwikkeling binnen de gedrags- en maatschappijwetenschappen: een internationale verkenning van ontwikkelingen en knelpunten. In: Symposium Statistische Software, Amsterdam 1985, p. 241-261.

P.G. Swanborn, Methoden van sociaal-wetenschappelijk onderzoek. Meppel 1981.

Ontvangen: 15-09-1987
Geaccepteerd: 08-03-1988