

University of Groningen

De computer binnen de sociale wetenschappen

Popping, R.; Stokman, F.N.

Published in:
EPRINTS-BOOK-TITLE

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:
1991

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

Citation for published version (APA):
Popping, R., & Stokman, F. N. (1991). De computer binnen de sociale wetenschappen: Een overzicht. In *EPRINTS-BOOK-TITLE* Bohn, Stafleu, Van Loghum.

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.

De computer binnen de sociale wetenschappen. Een overzicht

R. Popping en F.N. Stokman

1 Inleiding

Nieuwe computertoepassingen zijn aan de orde van dag, waarvan er vele ook van belang voor de sociale wetenschappen zijn. Was computergebruik daar eerst beperkt tot rekenwerk en tekstverwerking, nu worden computers binnen alle fasen van het onderzoek op zeer verschillende wijzen toegepast. Hieronder geven wij een overzicht daarvan, met name binnen sociaal-wetenschappelijk onderzoek. Daartoe zullen wij een tiental fasen in het wetenschappelijk onderzoek onderscheiden die aansluiten bij het verloop van het onderzoek zoals we dat doorgaans tegenkomen. Alvorens op deze fasen in te gaan, zal een aantal opmerkingen worden gemaakt met betrekking tot de ontwikkelingen in het gebruik van de computer zoals die zich tot op heden hebben voorgedaan.

2 Ontwikkelingen in het gebruik van de computer

De eerste computers kwamen beschikbaar in de tweede helft van de veertiger jaren. Deze computers werden gebruikt voor elementaire rekenoperaties. Spoedig volgden complexere rekenoperaties, simulaties en heuristieken. Om te rekenen wordt gebruik gemaakt van rekenvoorschriften of algoritmen. Bij simulatie wordt een model van een systeem gemaakt dat zo goed mogelijk de werking van het oorspronkelijke systeem benadert. Met behulp van programma's wordt nu getracht het gedrag van dat systeem zo goed mogelijk te voorspellen. Bij heuristieken maakt men gebruik van interpretaties en taxaties die verkregen zijn op grond van inzicht en ervaring. Zij worden gebruikt wanneer algoritmen niet toereikend zijn. Vervolgens kwam de data informatie retrieval, technieken voor het opslaan en terugvinden van data, doorgaans in de vorm van teksten. Hierdoor werd het mogelijk grote databestanden te raadplegen en met de uitkomsten daarvan te werken. Hieruit zijn de moderne *database mana-*

gement systemen voortgekomen. Los van deze ontwikkelingen ontstond het idee om de computer niet alleen te gebruiken voor het manipuleren van getallen, maar ook voor het omgaan met kennis. Dit leidde tot de zogeheten *kennissystemen*, waarin opslag van kennis en probleemoplossingsmethoden centraal staan.

De ontwikkeling in het gebruik van de computer binnen het empirisch sociaal-wetenschappelijk onderzoek volgt deze ontwikkeling op de voet. Ook hier wordt de computer in eerste instantie als *rekenautomaat* gebruikt. Als gesproken wordt over rekenen, dan hebben we het ook meteen over opslag en verwerking van grote hoeveelheden data. Opslag, manipulatie en documentatie van data is dan ook de volgende, direct met het rekenen verbonden taak van de computer binnen het onderzoek. Aangezien veel empirische onderzoekers geen statisticus zijn en ook geen hogere programmeertalen kennen, groeide snel de behoefte aan *standaardisering* van de beschikbare statistische programmatuur in geïntegreerde pakketten, waarin opslag, documentatie, procedures voor manipulatie en analyse van gegevens toegankelijk worden gemaakt voor niet-statistici. Hiermee ontstaat wel het risico van verkeerde toepassingen. Met betrekking tot dergelijke programmatuur stelt Van Reeken dat geen enkel rekencentrum genoeg statistici zal hebben "... to screen the statistical computations carried out on their machine for statistical misapplications. .. I cannot imagine that statisticians are the only ones using these programs. The only possible solution I see for this problem is to have 'fool-proof' programs." (Van Reeken, 1970: 74). Dergelijke "gekbestendige" programma's moeten ook nog "algemeen" zijn, dat wil zeggen alle standaard statistische technieken bevatten. "Gekbestendige" programma's zijn nooit gerealiseerd, maar "algemene" programma's, waarin data-opslag, -manipulatie, -documentatie, en standaard analyses zijn geïntegreerd, kwamen er heel snel. Denk aan pakketten als SPSS (Statistical Package for the Social Sciences), BMDP (BioMeDical Programs) en later SAS (Statistical Analysis System). Daarnaast bleven aparte programma's beschikbaar voor specifieke problemen en voor recent ontwikkelde modellen en technieken, zoals de beproeving van meetschalen en de toetsing van wiskundige modellen. Binnen Nederland is met name op dit terrein een dynamisch beleid gevoerd (zie Stokman, 1983^b)¹.

Na de rekentoepassingen komen binnen de sociale wetenschappen nieuwe computertoepassingen dankzij de inschakeling van de computer bij data informatie retrieval en het oplossen van problemen waarvoor geen analytische algoritmen beschikbaar zijn. De *data retrieval* is vooral belangrijk voor onderzoek waarbij inhoudsanalyse wordt gebruikt. In het begin is hier door de Nederlandse sociaal-wetenschappers nauwelijks

gebruik van gemaakt (Gadourek, 1985: 99). Voor het oplossen van problemen waarvoor geen analytische algoritmen beschikbaar zijn, wordt in toenemende mate *simulatie* toegepast, eerst bij onoplosbare numerieke problemen, daarna, mede dankzij heuristiek, ook bij meer algemene problemen. Hiermee is tevens een belangrijke stap gezet in de richting van *beslissing-ondersteunende systemen* en zogenaamde *kennissystemen*, waarin opslag van kennis en probleem-oplossingsmethoden centraal staan. De computer kan uit gegevens (de opgeslagen kennis) "conclusies trekken" en complexe vragen "beantwoorden", door uit de aanwezige alternatieven het beste te kiezen². Deze nieuwe toepassingen maken het mogelijk de computer te gebruiken voor de verdere uitbouw van complexe theorieën en de bestudering van de implicaties daarvan.

De toepassingsmogelijkheden van de computer zijn in de loop der tijd dus al drastisch verruimd. Verder heeft ook de massale introductie van de personal computer (PC) sterk aan deze verruiming bijgedragen. PC's bieden een aantal voordelen boven centrale computers. Zo is men niet meer afhankelijk van plaatsgebonden mainframes; de PC is gemakkelijk te transporteren en kan tegenwoordig ook vrij grote opdrachten aan (zie voor verdere voordelen Nagel, 1986).

Belangrijker is echter dat het met name de PC is geweest die de associatie tussen computer en rekenen in de traditionele betekenis van het woord radicaal heeft veranderd. De PC wordt een integraal onderdeel van alle bureauwerkzaamheden van een onderzoeker, primair als tekstverwerker en als administratief hulpmiddel, maar door de snelle toename van haar rekencapaciteiten en beschikbare programmatuur ook in toenemende mate als het medium voor data-opslag, dataverwerking en eigen rekenwerk. Bovendien hebben juist de krachtige versies van de PC's, de zogenaamde werkstations, een fundamentele bijdrage geleverd om de associatie computer - rekenmachine te ontkoppelen en de hierna te behandelen geheel nieuwe computertoepassingen binnen onderzoek mogelijk te maken.

Door de belangrijke plaats van de PC is de kracht van de centrale computer of computers en het aantal interactieve toegangen daartoe niet meer van primaire betekenis voor geavanceerd en grootschalig sociaal-wetenschappelijk onderzoek. De betekenis daarvan is nog verder afgenomen door de ontwikkeling van computer netwerken tussen de universiteiten, in Nederland met het SURF-net en op Europees niveau met het EARN-net, dat zelf weer verbonden is met andere belangrijke netwerken als BITNET³. Deze netwerken betekenen de mogelijkheid tot snelle communicatie met en toegang tot andere computers, waarbij SURF-net

in principe alle universitaire computerfaciliteiten openstelt voor alle Nederlandse onderzoekers.

3 Onderzoeksfasen

Binnen empirisch sociaal-wetenschappelijk onderzoek kunnen diverse fasen onderscheiden worden. De meest bekende is de empirische cyclus van De Groot (1961). Hieronder geven we een indeling in 10 fasen. Deze indeling sluit aan bij bovengenoemde cyclus en is mede gebaseerd op een indeling van De Boer (1983)⁴. Voor elke fase zal worden aangegeven hoe computers daarbinnen gebruikt kunnen worden. Vaak gebeurt dit al, soms liggen te behandelen toepassingen nog in het verschiet. De interpretaties "informatica" en "informatiekunde", zoals in de eerste bijdrage onderscheiden, lopen hierbij door elkaar. De fasen zijn:

1. probleemverkenning;
2. probleemformulering / modelspecificatie;
3. specificatie van de onderzoeksopzet;
4. keuze statistische techniek;
5. verzameling gegevens;
6. opslag en documentatie gegevens;
7. uitvoering statistische berekeningen;
8. interpretatie uitkomsten;
9. presentatie gegevens;
10. implementatie resultaten.

Het gebruik van de computer in de sociale wetenschappen is bij fase 7 begonnen en heeft zich van daaruit verspreid naar andere fasen (zie ook De Boer, 1983). Hiermee is een enorme diversificatie van het computergebruik opgetreden. In de volgende paragraaf wordt dit per fase uitgewerkt. Op het niveau van software en hardware zien wij echter tegelijkertijd een tendens tot integratie van deze toepassingen binnen een beperkt aantal "computeromgevingen" waarbinnen een onderzoeker werkt. Dit laatste aspect mag door de uitsplitsing niet uit het oog worden verloren.

4 Computertoepassingen binnen onderzoeksfasen

4.1 *Probleemverkenning*

In de probleemverkenningfase is de onderzoeker vooral bezig met het verzamelen, lezen en registreren van literatuur. In deze fase is de computer in eerste instantie te gebruiken bij het zoeken en met name bij de

geautomatiseerde literatuurontsluiting. "Computer searches" in de voor het probleem relevante internationale publikatie-bestanden via netwerken en cd-rom-systemen zijn nu onmisbaar geworden om een goede aansluiting van het onderzoek op de internationale literatuur te verkrijgen⁵. Datzelfde geldt voor de consultatie van de catalogus van de eigen (universiteits-) bibliotheek, waarbij boeken kunnen worden opgevraagd, nageetrokken kan worden of boeken al dan niet uitgeleend zijn en boeken te leen of ter inzage kunnen worden gereserveerd. De relevante publikaties zullen in het algemeen kunnen worden opgeslagen in een eigen flexibele database, waarin aantekeningen over de publikatie later kunnen worden toegevoegd en waaruit tenslotte de referenties voor publikaties weer kunnen worden gegenereerd. Indien vanuit de computer searches en de consultatie van de (universiteits-) bibliotheek titelbeschrijvingen op een elektronisch medium beschikbaar zijn dan kunnen ze zo zonder overschrijfwerk aan de eigen database worden toegevoegd. Naast de internationale literatuurbestanden is in Nederland een dergelijk bestand ontwikkeld op het gebied van de methodologie (SRM, 1990).

Verder kan de computer in deze fase gebruikt worden als ideeënverwerker. Dergelijke programma's komen in paragraaf 4.9 aan de orde.

4.2 Probleemformulering / modelspecificatie

Op basis van de probleemverkenning zal in het algemeen een eerste probleemaafbakening en -formulering plaatsvinden in de vorm van een onderzoeksvoorstel en -begroting. Het opstellen van het voorstel zal via een tekstverwerker gebeuren, voor de begroting kan een *spreadsheet* programma worden gebruikt. Een dergelijk programma is gebaseerd op een tweedimensionale matrix. De cellen in deze matrix kunnen drie soorten informatie bevatten: teksten, getallen en formules. Bij de formules ziet men de uitkomst op het scherm, niet de formule zelf. De formules bevatten doorgaans verwijzingen naar andere cellen, die getallen bevatten. Als waarden van getallen worden gewijzigd, kan men meteen zien hoe dit doorwerkt in de uitkomsten van de formules. De resultaten van alternatieve begrotingsberekeningen binnen een spreadsheet kunnen direct binnen het eigen tekstverwerkingsprogramma worden gehaald. De uitgewerkte tekst kan nu reeds veelal via het netwerk bij het secretariaat worden gedeponerd, waar de lay-out verder wordt vervolmaakt en standaard gegevens worden toegevoegd, waarna het manuscript via het netwerk bij een der reproductie-afdelingen ter vermenigvuldiging wordt gedeponerd.

Voor de modelspecificatie zijn zowel statische modellen als procesmodellen van belang. Tot voor kort lag bij onderzoek de meeste nadruk op

wat we statische modellen zullen noemen. Dit zijn modellen waarmee het mogelijk is uitspraken te doen over de empirie op een bepaald moment. Ook kunnen ze gebruikt worden om het verschil tussen twee of meerdere momenten vast te stellen. Dit soort modellen wordt momenteel nog doorgaans gebruikt bij het sociaal-wetenschappelijk onderzoek. De resultaten die gevonden worden, kunnen betrekking hebben op de uitkomsten van sociale processen, ze geven geen informatie over deze processen zelf. Momenteel komt er daarnaast steeds meer aandacht voor onderzoek waarbij men wel kennis wil krijgen over de processen zelf. We hebben dan te maken met modellen waarin wordt geprobeerd tot een goede representatie te komen van sociale processen die zich binnen de empirie afspelen. Dit betreft zowel de structuur als de functionaliteit van deze processen. Dit noemen we procesmodellen. Vaak wordt hierbij gebruik gemaakt van computersimulaties.

4.2.1 Statische modellen

Statische modellen spelen een belangrijke rol in de vierde fase (de keuze van statistische technieken), maar moeten al goed doordacht worden bij de modelspecificatie. De introductie van de modelbouw heeft zorg gedragen voor belangrijke veranderingen binnen de sociale wetenschappen. Modellenbouw en modellentoetsing moeten het toetsen van fundamentele hypothesen vervangen (Gadourek, 1985: 97). Voor de toetsing is veel rekenwerk nodig, vooral daar waar moet worden gekomen tot schattingen van modellen. Hierbij is de computer onontbeerlijk. Veel modellen die gebruikt worden behoren tot de categorie van statische modellen.

Dit heeft zowel betrekking op meet- als analysemodellen. Toepassing van meetmodellen geeft een grotere validiteit aan metingen, omdat met een meetmodel veronderstellingen omtrent de meting daadwerkelijk worden getoetst terwijl die bij ad-hoc meting zonder meer worden aangenomen of niet worden geëxpliciteerd. Wanneer wij bijvoorbeeld veronderstellen dat een vragenbatterij betrekking heeft op een latent kenmerk (zoals een bepaalde vaardigheid), dan stelt een meetmodel ons in staat om na te gaan of "moeilijke" vragen ook voornamelijk goed worden beantwoord door mensen die ook de gemakkelijkere vragen goed hebben beantwoord. Stokman (1983^a) heeft een aantal kenmerken van het meetmodel gespecificeerd:

- het specificeert hoe empirische relaties, ook manifest gedrag genoemd, gerelateerd zijn aan latente kenmerken;
- het omvat een schattingstheorie op basis waarvan onbekende groot-heden kunnen worden geschat;

- het omvat een toetsingstheorie op basis waarvan een model getoetst kan worden aan de hand van de restricties die uit het model voor het manifeste gedrag zijn af te leiden.

In structurele modellen worden aard en richting van samenhang tussen verschillende (latente) kenmerken gespecificeerd, waardoor de structuur tussen de verschijnselen wordt vastgelegd. Bij meetmodellen moeten drie belangrijke keuzes worden gemaakt, die in de modelspecificatie terug moeten komen:

1. de keuze van objecten als drager van kenmerken;
2. de keuze van indicatoren die geacht mogen worden betrekking te hebben op het te meten latente kenmerk;
3. de keuze van de vast te leggen empirische relaties tussen objecten en indicatoren.

Uit deze specificaties kunnen restricties worden afgeleid waaraan de gegevens moeten voldoen en die voor de toetsing van het model kunnen worden gebruikt. Zijn we geïnteresseerd in de verschillen die zijn opgetreden tussen twee meetmomenten, dan moeten we van tevoren vastleggen wanneer we mogen spreken over een verschil, wat moet er dan veranderd zijn.

In een structureel of analysemodel wordt aard en richting van samenhang tussen verschillende (latente) kenmerken gespecificeerd en wordt daarmee de structuur tussen die verschijnselen vastgelegd.

Dit overdenken van de modelspecificatie geldt met name voor experimenteel onderzoek, waar de specificatie direct met het ontwerp van het experiment is geïntegreerd. Het programma MEL (Micro Experimental Laboratory) (Schneider, 1989) is een goed voorbeeld van de nieuwe mogelijkheden die computers voor onderzoek gebaseerd op statische modellen bieden. Met behulp van dit programma is het mogelijk een experiment te ontwerpen, vragenlijsten te ontwerpen en af te nemen en tenslotte via het programma ook de data-analyse te verzorgen. Een ander voorbeeld is APLEX, een programmeertaal die gemaakt is voor experimentele opstellingen en testafnames waarbij een nauwkeurige tijdmeting gevraagd wordt en waarbij een flexibele afhandeling van grafische en/of tekstuele stimuli van belang is (Tibosch, 1989). De computer stuurt het experiment en registreert de reacties. Het is zelfs mogelijk een en ander zo te programmeren dat de stimuli in random volgorde worden aangeboden, dit om volgorde effecten te vermijden.

Het moge duidelijk zijn dat de hier genoemde programma's ook betrekking hebben op de specificatie van de onderzoeksopzet.

4.2.2 Procesmodellen

Bij de procesmodellen komen we dankzij de computer meer nieuwe mogelijkheden tegen. In de fase van de definitieve probleemformulering en modelspecificatie kunnen computersimulaties en in ontwikkeling zijnde kennisrepresentatiesystemen een belangrijk hulpmiddel zijn om tot een betere modelspecificatie te komen dan daar zonder mogelijk is. Eerdere simulatiemodellen⁶ van maatschappelijke verschijnselen, zoals de macro-modellen van de Club van Rome, bestonden uit een stelsel van wiskundige vergelijkingen, waarin de ontwikkeling van verschillende macroverschijnselen op elkaar worden betrokken.

Nu realiseert men zich in toenemende mate dat een goede verklaring van maatschappelijke verschijnselen slechts mogelijk is wanneer in die verklaring het menselijk keuzegedrag als essentieel element is opgenomen (zie onder andere Boudon & Bourricaud, 1982; Coleman, 1986; 1990). Dit gedrag is zelf weer afhankelijk van, *geconditioneerd door* de maatschappelijke context waarbinnen men zich bevindt. Deze context is namelijk medebepalend voor de beschikbare of gepercipieerde gedragsalternatieven en de daaraan verbonden kosten en baten, niet alleen in materiële zin, maar ook in termen van sociale waardering (Lindenberg, 1989). Maatschappelijke verschijnselen ontstaan dan ook onder invloed van gelijktijdig handelende, deels op elkaar reagerende actoren, die op basis van herhaaldelijke evaluaties hun gedrag optimaliseren en op basis daarvan al dan niet aanpassen⁷. Een eenvoudig voorbeeld hiervan geeft Schelling (1978). Hij laat zien hoe onder bepaalde eenvoudige aannames segregatie in woonwijken kan resulteren vanuit individuele keuzes die "whites" en "blacks" maken om naar of uit woonwijken te verhuizen op grond van een voor hen maximaal getolereerde ratio tussen de twee rassen in een woonwijk. Hier komt duidelijk tot uitdrukking hoe het gedrag van individuen bepaald wordt door de keuzen die anderen maken. Het gaat derhalve om parallel, onder verschillende restricties opererende, zelflerende en deels op elkaar reagerende actoren, waaruit maatschappelijke verschijnselen ontstaan. De uitgangspunten van object-georiënteerde modellering sluiten hier naadloos op aan. Het basisidee achter object-georiënteerde modellering betreft met elkaar communicerende objecten, waarbij elke wijziging in zijn toestand slechts door het object zelf kan worden aangebracht. Deze wijzigingen kunnen resulteren in boodschappen aan andere objecten die vervolgens op basis hiervan tot handelingen kunnen worden aangezet. Het principe van het zogeheten polymorphisme maakt het daarbij mogelijk dat verschillende objecten op dezelfde boodschap op hun eigen wijze reageren, terwijl het principe van de overerving van eigenschappen en methoden de mogelijkheid biedt tot een

overzichtelijke structuur tussen objecten en een overzichtelijke ordening van methoden (Pinson & Wiener, 1988). Object-georiënteerde programmering is het middel om tot een adequate representatie te komen en modellen door te rekenen op hun implicaties en tegenstrijdigheden. Het zijn juist de moderne grafische werkstations die niet alleen de benodigde hardware faciliteiten, maar vooral ook de software faciliteiten bevatten, die dit soort simulaties mogelijk maken, mede omdat zij reeds kant en klare toegankelijke toepassing programma's bevatten die gemakkelijk aan de eigen vraagstelling zijn aan te passen. Bovendien beschikken zij over zeer sterke grafische faciliteiten, die de resultaten van de simulatie zowel naar de onderzoeker als in het eindrapport inzichtelijk maken.

Omdat het gaat om parallel opererende actoren zijn ook de ontwikkelingen op het terrein van parallel distributing processes uitermate relevant. Bij een computer worden alle operaties nu nog na elkaar afgehandeld via één processor, bij de bedoelde processen is het de bedoeling dat ze naast elkaar kunnen worden afgehandeld, liefst ook nog onafhankelijk van elkaar. Dit is vooral relevant wanneer er flexibele mogelijkheden zijn tot wisselende terugkoppeling tussen processoren, waarbij deze terugkoppeling mede afhankelijk is van de uitkomsten van het gedrag (de uitkomst). Een dergelijk flexibel systeem vinden wij bijvoorbeeld in de Connections Machine, waarin 65.536 kleine processoren parallel opereren en, voorzover gewenst, met elkaar kunnen communiceren (Hillis, 1987).

Het is frappant hoe bovenstaande ontwikkelingen parallel lopen aan ideeën rond neurale netwerken als zichzelf organiserende systemen, waarbij neuronen andere neuronen activeren cq. deactiveren en hoe hieruit de drie fundamentele processen van "extinction, recognance and the formation of connections" kunnen worden afgeleid (Rumelhart & McClelland, 1982; De Vries, 1989).

Het opstellen van modellen is niet zonder problemen. In het artikel van Vennix wordt een methode beschreven die gebruikt kan worden bij het bouwen van modellen.

Een andere vorm van simulatie die zeker genoemd moet worden is de spelsimulatie, ook gaming genoemd. Hierbij wordt het gedrag van een groep personen nagebootst. De deelnemers beschikken over bepaalde informatie en zijn gebonden aan spelregels. Door de computer er bij te betrekken, kan de proefleider van buiten informatie geven aan de deelnemers zodat de deelnemers niet meer alleen op elkaar reageren. Dit wordt ook interactieve simulatie genoemd. Deze vorm van simulatie is van belang daar waar echte experimenten niet mogelijk zijn.

Samenvattend, zijn relaties tussen variabelen kwantificeerbaar dan hebben we te doen met numerieke simulatie. Hiervoor kunnen numerieke oplossingen gegenereerd worden. Als kwantificering niet mogelijk is dan dienen heuristische oplossingen gezocht te worden. Dit kunnen ALS-DAN-beslissingen zijn. Binnen sociaal-wetenschappelijke domeinen hebben we vaak met een dergelijke situatie te maken. Hier hebben we dan al te doen met een kennisrepresentatiesysteem. Indien simulatie wordt gebruikt voor theorieconstructie, dan informeert de uitkomst van de simulatie over de waarschijnlijkheid van de gemaakte aannames⁸.

4.3 Specificatie van de onderzoeksopzet

Bij de specificatie van de onderzoeksopzet kan het uitermate wenselijk zijn eerst via databanken na te gaan of de benodigde gegevens reeds in eerder onderzoek zijn vergaard om deze vervolgens via secundaire analyse vanuit de eigen probleemstelling te analyseren. Helaas is het nog niet mogelijk het Nederlandse data-archief voor sociaal-wetenschappelijk onderzoek te Amsterdam, het Steinmetz Archief, interactief te consulteren. Wel is het al mogelijk bij een databank voor sociaal-wetenschappelijke gegevens in Keulen, het Zentral Archiv. Voor een deel van hun bestanden is het zelfs mogelijk om via het netwerk onmiddellijk selecties uit bestanden te maken en de relevante gegevens voor analyse via het netwerk beschikbaar te krijgen.

Zoals hiervoor al is aangegeven, staan aan de basis van meetmodellen stimuli of items. Het is mogelijk grote verzamelingen items op te stellen waar meetmodellen uit geselecteerd kunnen worden. Zo'n verzameling items heet een itembank. Momenteel wordt gewerkt aan het ontwikkelen van procedures voor het opzetten van dergelijke itembanken en het construeren van toetsen uit dergelijke banken. Bij de itemselectie gaat het daar om het vinden van de beste items gegeven tal van praktische voorwaarden. Procedures om dergelijke tests op te bouwen worden behandeld door Boekkooi-Timminga (1989). Het opstellen van een test kan een gigantische opgave zijn. Op grond van 10 items zijn al $2^{10} = 1024$ tests mogelijk.

De computer kan ook van groot nut zijn wanneer in een onderzoek heel specifieke respondenten met elkaar vergeleken moeten worden. Zo is het achteraf of "ex-post-facto" experiment nu gemakkelijker te gebruiken omdat de computer gebruikt kan worden bij het "matchen", het selecteren van paren van respondenten die, afgezien van het te onderzoeken kenmerk, gelijkgeschakeld zijn.

Aangezien een verklaring van maatschappelijke verschijnselen, zoals hiervoor gesteld, gedragsgegevens van individuen in combinatie met

gegevens over collectieve verschijnselen behoeft, zullen datasets in de sociale wetenschappen in toenemende mate een complex karakter hebben. In deze datasets zullen gegevens op verschillende maatschappelijke niveaus aanwezig zijn en op elkaar moeten worden betrokken. Het kan reeds in deze fase gewenst zijn via prototyping een database uit te testen om na te gaan of alle gewenste relaties ook daadwerkelijk kunnen worden gerealiseerd en -indien relevant- gebruikersvriendelijke schermen voor de invoer van gegevens te definiëren. Prototyping is het tonen van een toekomstig systeem of deel ervan in een op echt gelijkende afbeelding. Als voorbeeld kan de uitwerking van een relationele database worden genoemd om de gegevens op te slaan waarop een geïntegreerd besluitvormingsmodel is gebaseerd. In dit model wordt expliciet onderscheid gemaakt in twee fasen van de besluitvorming, de fase van de uiteindelijke besluitvorming waarin beslismacht essentieel is, en de daaraan voorafgaande fase waarin het gaat om invloed bij de totstandkoming van het beleid (Stokman & Van den Bos, 1991; Van den Bos, 1991). Met behulp van het model is het mogelijk de uitkomst van beslissingen te voorspellen. Tevens kan met het model worden berekend welke organisaties als machtigste dienen te worden beschouwd, zowel in het totale besluitvormingsproces als in elk van de fasen apart. Het model vereist gegevens over toegangsrelaties tussen actoren, machtsbronnen van actoren, alsmede gegevens over hun stellingname ten aanzien van, hun beslismacht over en hun belang bij beleidsbeslissingen. In het prototype werden alle belangrijke tabellen gedefinieerd en werd nagegaan of alle berekeningen voor de voorspelling van de uitkomsten van de beslissingen en de bepaling van de macht van de actoren konden worden uitgevoerd. In een eerste toepassing bleek het mogelijk om van een achttal belangrijke en controversiële beslissingen in het Amerikaanse Congres op het gebied van energiepolitiek er zeven goed te voorspellen.

Met de verdergaande informatisering van de maatschappij zullen in toenemende mate gegevens voor sociaal-wetenschappelijk onderzoek beschikbaar komen die zelf al op een elektronische informatiedrager staan. Veelal gaat het hier om gigantische bestanden, waaruit voor het onderzoek relevante gegevens moeten worden geselecteerd. Dit geldt niet alleen voor sociaal-wetenschappelijk onderzoek, maar ook in toenemende mate voor historisch onderzoek⁹.

4.4 Keuze statistische techniek

Met betrekking tot de keuze van statistische techniek zijn er ook veel meer mogelijkheden ontstaan, vooral als gevolg van het feit dat heel complexe rekenpartijen nu heel snel door de computer kunnen worden gedaan. De keuzen die men heeft worden mede bepaald door de modelspecificatie (zie vooral paragraaf 4.2.1).

De vernieuwingen hebben vooral betrekking op een drietal zaken, te weten het meetniveau, het toetsen van gehele modellen en de exploratie van data.

Bij veel sociale kenmerken kan slechts aangegeven worden dat zij gelijk aan of verschillend van elkaar zijn (bijvoorbeeld jouw en mijn godsdienst; we spreken dan van een nominaal meetniveau), bij andere is hoogstens een ordening aan te brengen in termen van meer of minder, maar niet hoeveel meer of minder (ordinaal meetniveau). Ligt het hoeveel meer of minder wel vast dan spreken we over interval niveau. Betekent een drie maal zo hoge meetwaarde ook dat het kwantitatieve kenmerk echt drie maal zo vaak voorkomt, dan spreken we over ratio niveau (als voorbeeld: Jan heeft drie maal zoveel appels als Piet). Moderne technieken in de sociale wetenschappen houden daar rekening mee. Soms kan dit meetniveau via meetmodellen worden verhoogd, waarbij die extra informatie via veronderstellingen in het meetmodel wordt "gekocht". Vanzelfsprekend moeten die assumpties dan wel worden getoetst. Zo kunnen via toepassing van het zogenaamde Rasch model antwoorden op een batterij "ja/nee vragen" tot een interval score van de moeilijkheidsgraad van de vragen en tot een vaardigheidsscore voor de respondenten worden omgevormd.

Daarnaast kunnen thans *gehele structurele modellen* getoetst worden in plaats van alleen verbanden tussen variabelen. Bekende voorbeelden daarvan zijn het zogeheten LISREL model (Linear Structural Relationship) en de lineaire modellen in het algemeen. Logistisch en log-lineaire modellen stellen de onderzoeker in staat dit ook te doen bij een laag meetniveau.

Voor tal van meet en analyse modellen geldt dat zij zeer rekenintensief zijn, vooral als iteratie nodig is om te komen tot een beste schatting van een model. Het gebruik van dergelijke modellen is alleen mogelijk dankzij de computer.

Tenslotte noemen we de *exploratieve data analyse*. Er zijn allerlei hulpmiddelen ontstaan met behulp waarvan men nu in staat is beter en sneller inzicht te krijgen in het datamateriaal waar men over beschikt alvorens met het echte analyseren te beginnen. Deze vorm van data analyse was oorspronkelijk bedoeld om met de hand uitgevoerd te

worden. Met behulp van de computer kan het rekenwerk echter heel snel gebeuren en kunnen de resultaten op heel overzichtelijke wijze op het beeldscherm gerepresenteerd worden.

Hier willen we ook nog noemen dat er dankzij de computer meer inzicht kan worden verkregen in het gedrag van allerlei statistische indices en zelfs modellen. Dit gebeurt in simulaties van stochastische processen of "Monte Carlo" studies. Zo verrichtte Popping (1984) een Monte Carlo studie met betrekking tot een aantal typen overeenstemmingsmaten en verrichtte Boomsma (1983) een dergelijke studie met betrekking tot de vraag of toepassing van het LISREL model leidt tot goede schattingen wanneer aan een aantal assumpties van het model niet is voldaan.

4.5 Verzameling gegevens

Computers spelen een toenemend belangrijke rol bij de verzameling van sociaal-wetenschappelijke gegevens. Telefonische enquetering vindt geheel gecomputeriseerd plaats via zogeheten CATI-systemen (Computer Assisted Telephonic Interviewing). De vragenlijst is opgeslagen in de computer, de antwoorden worden onmiddellijk ingetypt. Vaak hoeven blokken van vragen slechts gesteld te worden als respondenten bepaalde eigenschappen hebben. Zo hoeven specifieke vragen over het beroep slechts gesteld te worden aan respondenten die werken. Bij CATI-systemen wordt deze zogenaamde *routing* door de vragenlijst eveneens in de computer vastgelegd, zodat geen onnodige of zinloze vragen worden gesteld. Zelfs de selectie van het telefoonnummer en het opbellen zelf kunnen computergestuurd zijn. De gegevens worden rechtstreeks toegevoegd aan een database die geschikt is voor statistische analyses. Het belang van ontwikkelingen rond de "slimme" telefoon voor dit soort onderzoek behoeft nauwelijks betoog.

Computergestuurde vraaggesprekken vinden momenteel ook plaats bij persoonlijke interviews, waarbij gebruik gemaakt wordt van *hand-held* computers, waarin de vragenlijst is opgeslagen. De belangrijkste grote enquêtes van het CBS, zoals de telling van de beroepsbevolking (EBB) en het periodieke woningonderzoek worden momenteel zo afgenomen. Alle fasen van codering en invoer van de gegevens, met alle mogelijke foutenbronnen zijn op deze wijze uitgeschakeld. De gegevens die overdag of 's avonds door de interviewers op de hand-help computer verzameld zijn kunnen 's nachts automatisch via een modem overgebracht worden naar de centrale computer van de instelling die met het afnemen van de vragenlijsten belast is. Voor een uitgebreider overzicht betreffende de

mogelijkheden bij deze vorm van gegevensverzameling verwijzen we naar Sikkel (1987) en naar de bijdrage van Bethlehem in de bundel.

Uiteraard moet duidelijk worden in hoeverre deze nieuwe vormen van vragenlijstafname van invloed zijn op de betrouwbaarheid en validiteit van de verkregen data. Nederhof (1987) geeft een eerste overzicht van onderzoeken met betrekkingen tot dit probleem. Tal van vragen zijn evenwel nog niet beantwoord.

Een ander soort datavergaring dat we willen noemen betreft de afname van psychologische tests. Ook deze afname verloopt in toenemende mate via de computer. Zo kunnen individuele verschillen met betrekking tot sociale waarden via een computerprogramma gemeten worden (Liebrand & McClintock, 1988).

Video-opnamen en computergestuurde codering van observatiegegevens (Hendersen, 1989) en computergestuurde experimenten zijn andere voorbeelden waaruit blijkt dat computers niet alleen meer een rol spelen voor de analoge invoer van fysische reactiegegevens, die kenmerkend waren voor de vroegere experimenten die interactief werden uitgevoerd (zie ook paragraaf 4.2.1).

Tenslotte zal naar verwachting de inhoudsanalyse van documenten, waarvan de methodologie in de jaren zeventig stakte wegens haar zeer arbeidsintensieve karakter, in de wat verdere toekomst volop kunnen profiteren van de ontwikkelingen op het gebied van de taalanalyse. Momenteel zijn er hulpmiddelen met betrekking tot het toewijzen van tekstfragmenten waarbij rekening kan worden gehouden met de socio-culturele omgeving (Carley, 1988) en worden er analysemethoden ontwikkeld waarbij gebruik wordt gemaakt van grafen analyse (Van Cuilenburg e.a. 1988). Voor een overzicht van de betekenis van de computer voor allerlei vormen van kwalitatief onderzoek wordt verwezen naar Pfaffenberg (1988). De programmatuur voor zowel het toewijzen van tekstfragmenten als voor het opbouwen van de punten in een graaf is gebaseerd op ideeën uit de kunstmatige intelligentie. Gelet hierop en op zaken die in paragraaf 4.2.2 aan de orde zijn geweest, onderschrijven we van harte de verzuchting van Anderson:

"It is time for sociology to break its intellectual isolation and participate in the cognitivist rethinking of human action, and to avail itself on theoretical ideas, techniques and tools that have been developed in AI and cognitive science."
(Anderson, 1989)

4.6 Opslag en documentatie gegevens

Pas nu komen we toe aan de fasen waartoe tot het begin van de jaren tachtig de computertoepassingen vrijwel beperkt waren: de opslag en documentatie van gegevens en de uitvoering van de statistische berekeningen. Werden wij vroeger gedwongen onze gegevens te persen in een rechthoekige datamatrix, nu is het mogelijk een sociaal systeem af te beelden op een relationele database. Dit is een database waarin alle (unieke) informatie opgeslagen is in tabellen die men relaties noemt. Elke tabel bevat een of meerdere kolommen die informatie bevatten via welke het mogelijk is te verwijzen naar informatie in andere tabellen. Dit is de zogeheten sleutel. Nu is het mogelijk gegevens uit een complex terrein uiteen te leggen in een aantal enkelvoudige relaties waarin slechts een verband voorkomt, namelijk dat tussen de sleutel en het bijbehorende kenmerk. Omdat een bepaalde soort informatie slechts eenmaal voorkomt heeft men geen last van redundantie, elkaar tegensprekende gegevens. Gegevens uit verschillende tabellen kunnen via de sleutels aan elkaar gekoppeld worden zodat nieuwe (tijdelijke) tabellen ontstaan die dienst kunnen doen als datamatrix voor een statistische analyse. Gegevens uit tabellen waarop analyses verricht moeten worden, kunnen op verschillend aggregatieniveau zijn. Een tabel kan bijvoorbeeld betrekking hebben op de bewoners van een stad, een volgende op de wijken uit die stad. In deze laatste tabel kunnen dan wel gegevens opgenomen zijn die betrekking hebben op de bewoners zoals hun aantal of gemiddelde leeftijd (zie verder Janssens, 1987).

Een relationele database wordt gebruikt in Stokman e.a. (1989). Zij gebruiken deze database voor een onderzoek naar wetenschappelijke produktie. De database bevat 23 tabellen. Alle informatie is slechts een keer opgeslagen. In een rechthoekige database zou dit niet kunnen. Een tekst met drie auteurs zou driemaal opgenomen moeten worden, elke keer gerelateerd aan een andere auteur. Binnen het relationele model hoeft dat niet. Binnen dat model is het ook mogelijk analyses te verrichten op het niveau van de tekst (die wordt dan een keer geteld), maar ook op het niveau van de auteurs (nu kan ook nog vrij bepaald worden in welke mate de bijdrage van de auteur aan de tekst geteld wordt, bijvoorbeeld bij elke auteur volledig of slechts voor een derde deel).

Door Molenaar en Popping (1988) is over een experiment gerapporteerd waarin aan personen die nauwelijks kennis hadden met betrekking tot databases is gevraagd een database ten behoeve van een ledenadministratie te ontwerpen en te vullen met enige data. Bij deze taak doen zich twee hoofdgroepen van problemen voor. De eerste betreft de definitie van de database, de tweede betreft de implementatie. Vooral een

goede definitie van een database vereist training en inzicht, de implementatie is vaak op te lossen door de handleiding bij het pakket goed te bestuderen.

4.7 Uitvoering statistische berekeningen

Bij de keuze van de statistische techniek zal in het algemeen gebruik worden gemaakt van een bestaande techniek. Beschikbaarheid van gebruikersvriendelijke programmatuur is van groot belang. Veel statistisch geavanceerd onderzoek vindt echter plaats met specialistische technieken die niet in algemene statistische pakketten aanwezig zijn. Veelal zijn deze binnen universiteiten ontwikkeld. Het Interuniversitaire Expertisecentrum *ProGAMMA* te Groningen is er op gericht dergelijke producten algemener te verspreiden, maar ook de produktie zelf aanzienlijk te vereenvoudigen door via een strakke modulaire opbouw van programmatuur de gehele omgeving van een algoritme reeds ter beschikking te stellen, zodat de ontwikkeling van een nieuw programma vrijwel beperkt kan blijven tot de ontwikkeling van het algoritme zelf. Ook hierbij speelt een object-georiënteerde omgeving een belangrijke rol. In de nabije toekomst zijn dergelijke modulen via het netwerk consulteerbaar en opvraagbaar. Daarnaast geven moderne pakketten de mogelijkheid aan de gebruiker, bijvoorbeeld via de definitie van macro's, eigen modellen te specificeren. Voorbeelden van dergelijke pakketten zijn BMDP en GLIM (General Linear Interactive Modelling).

Hierboven is reeds gesteld dat nieuwe statistische modellen veel sterker dan vroeger rekening houden met het lage meetniveau van veel gegevens in de sociale wetenschappen. Daarnaast zijn ook speciale modellen ontwikkeld voor meer complexe gegevensstructuren. Een voorbeeld daarvan is de speciale methodologie voor sociale netwerkanalyse, waarin de structuur van dergelijke netwerken kan worden onderzocht naar de mate van hechtheid en gecentraliseerdheid, de positie van de punten kan worden bepaald naar diverse aspecten van centraliteit en de positie van punten onderling kan worden vergeleken op hun gelijkenis binnen het netwerk (Sprenger & Stokman, 1989). Een ander voorbeeld is de ontwikkeling van aangepaste lineaire modellen voor multilevel analyse, waarin eenheden en effecten op verschillende sociale niveaus gelijktijdig kunnen worden geanalyseerd.

Tenslotte geven interactieve analysefaciliteiten een veel directere terugkoppeling van de resultaten naar de gebruiker, zij het dat dit niet geheel onproblematisch is wegens kanskapitulatie en de neiging bij gebruikers om het logboek te verwaarlozen. Ze houden dan niet meer bij wat ze allemaal al uitgerekend hebben.

4.8 Interpretatie uitkomsten

Hoewel "gekbestendige" programma's nooit zullen bestaan, zijn wel de eerste kennissystemen in ontwikkeling die de onderzoeker kunnen helpen bij de keuze van de techniek en vooral de interpretatie van de uitkomsten. Visser en Slooff (1990) rapporteren over de kennisacquisitie ten behoeve van een kennissysteem dat gebruikt kan gaan worden bij het interpreteren van uitkomsten verkregen via correspondentieanalyse. Dit is een techniek waarmee het mogelijk is de relatie tussen rijen en kolommen in een tweedimensionale tabel visueel te maken in een meerdimensionale ruimte. Het blijkt dat hulp bij de interpretatie geboden kan worden door er zorg voor te dragen dat de resultaten op een overzichtelijke wijze worden gepresenteerd, waarbij uitsluitend die informatie wordt verstrekt die voor de interpretatie (en toetsing) van belang is. Verder zijn er een aantal regels geformuleerd waarin uitspraken worden gedaan met betrekking tot bepaalde te interpreteren gedeeltes. Dit kan ons inziens nog verder gaan: bij bepaalde uitkomsten dient een systeem een waarschuwing geven.

4.9 Presentatie gegevens

Voor de presentatie van de uitkomsten van het onderzoek zijn we weer terug bij de tekstverwerking.

Met betrekking tot de opbouw van een tekst zijn er ideeënverwerkers in omloop die hierbij als hulpmiddel kunnen fungeren. Programma's hiervoor zijn een soort tekstverwerkers welke beschikken over een zogeheten outline, een hiërarchisch opgebouwde inhoudsopgave. Het is hiermee mogelijk een tekst gemakkelijker op te bouwen. Met de inhoudsopgave kan gemanipuleerd worden. Men kan subniveau's maskeren, zodat alleen de hoofdlijn zichtbaar wordt, men kan zich op een niveau richten, men kan stukken tekst verplaatsen van een lager naar een hoger niveau of omgekeerd. Voor een verdere behandeling van mogelijkheden van dergelijke programma's wordt verwezen naar Heyting (1988). Gebruikers van dergelijke programma's blijken desondanks toch vaak moeite te hebben met het ordenen van gedachten en het uitschrijven van deze gedachten (Kellogg, 1986). Door gebruik te maken van de mogelijkheden op het gebied van desk-top publishing kan zorg worden gedragen voor een zeer verzorgde presentatie.

Database programma's hebben report faciliteiten waarmee het mogelijk is tabellen, zoals op te nemen in een tekst, te genereren. Deze tabellen en ook uitvoer van statistische programma's kunnen doorgaans zonder meer in teksten worden ingelast. Er zijn ook programma's (Hypercard) om

informatie uit verschillende bron afkomstig, bijvoorbeeld teksten, grafieken, te koppelen.

Ook is lang gedacht dat elektronisch publiceren erg belangrijk gaat worden. Dit gebeurt toch nog nauwelijks. De uiteindelijke publikaties verschijnen op papier. Alleen van handleidingen bij computer programma's verschijnt, naast de versie op papier, ook wel een versie op floppy disk. Een reden voor dit tegenvallen van elektronisch publiceren ligt bij de lezers. Bij een elektronisch gepubliceerde tekst is het lastig om snel een idee te krijgen van waar de tekst over gaat. Voor het bladeren, zoals bij tekst op papier kan, biedt de computer geen goede tegenhanger. Het is nu immers alleen mogelijk schermen te bekijken. Wel kan via de computer snel inzicht worden verkregen in de opbouw van een tekst (via de outline), maar dit weegt niet op tegen het bladeren.

4.10 Implementatie resultaten

Rond het jaar 2000 hopen we de inhoudelijke resultaten van onderzoek toe te kunnen voegen aan een geïntegreerde kennisbank waarin sociaal-wetenschappelijke kennis is opgeslagen die zowel voor beleids- als voor wetenschappelijke doeleinden kan worden geraadpleegd. Onderzoek naar kennisstructurering laat zien dat daarvoor nog wel enkele verdere ontwikkelingen nodig zijn.

5 Slot

In deze bijdrage hebben we een overzicht gegeven van ontwikkelingen in het gebruik van de computer binnen het sociaal-wetenschappelijk onderzoek. Daarbij is gebleken dat de associatie computer en rekenen verouderd is en los moet worden gelaten. Er is sprake van een grote mate van differentiatie. Om het leven van onderzoekers dragelijk te maken wordt door softwareleveranciers echter ook naar integratie gestreefd waardoor een onderzoeker zoveel mogelijk binnen een omgeving kan blijven werken. Dit laatste is voorlopig echter nog meer een ideaal dan werkelijkheid. Sociaal-wetenschappelijke onderzoekers hebben dan ook de moeilijke taak niet alleen op hun vakgebied bij te blijven maar ook op het gebied van computertoepassingen. Gezien de snelle ontwikkelingen in dit laatste is dat geen eenvoudige taak.

NOTEN

1. Evaluatie van standaardprogrammatuur is een aparte tak van onderzoek geworden. Het gaat daarbij met name om de eisen die op het gebied van gebruikersvriendelijkheid aan de programma's gesteld moeten worden. Voor een eerste overzicht en een voorbeeld van dergelijk onderzoek verwijzen we naar Francis (1983).
2. Kennissystemen worden tegenwoordig ook gebruikt om alternatieve meetinstrumenten optimaal te converteren (Popping, 1991).
3. Een netwerk is een hulpmiddel om computers met elkaar te verbinden. Dit kan gebeuren via kabels. Er wordt dan gesproken over een LAN (Local Area Network). Dit treffen we doorgaans aan als het computers betreft die in een gebouw staan opgesteld. Bij een LAN heeft men veelal oog voor het gezamenlijk gebruik van programmatuur, databestanden en printers. Staan de computers ver uitelkaar dan spreken we over een WAN (Wide Area Network). Hier worden de verbindingen gelegd via telefoonlijnen en satelieten. Deze netwerken worden veelal gebruikt voor het versturen van berichten en gegevensbestanden.

SURF-net (Samenwerkende Universitaire Rekencentra; nu is het een BV) is het landelijk communicatienetwerk voor instellingen van hoger onderwijs en onderzoek. Via dit netwerk is het zelfs mogelijk te werken op computers van andere instellingen. Het EARN-net (European Academic Research Network) is een computernetwerk dat computers van wetenschappelijke instellingen binnen Europa met elkaar verbindt. Dit netwerk is verbonden met een groter netwerk. Het Amerikaanse deel ervan heet BITNET. Via speciale computers (gateways) kunnen andere netwerken bereikt worden, als ARPANET, CONNET en JANET, de netwerken van universiteiten en onderzoeksinstellingen in de Verenigde Staten, Canada en Engeland. Het Nederlandse deel van EARN valt onder het SURF-net.
4. De Boer gebruikt de indeling vooral om het belang van inductieve analyse-technieken en 'gebruikersvriendelijkheid' van programmatuur aan te tonen.
5. Op korte termijn zal dit binnen de meeste universiteiten ook vanaf de werkplek mogelijk zijn. Met cd-rom wordt hier bedoeld op een literatuurbestand (referenties met samenvattingen) dat op een compact disk is opgeslagen en waar via een computer en enige speciale apparatuur (waaronder een cd-speler) in gezocht kan worden.
6. Om een indruk te geven van wat er zoal te simuleren valt het volgende. Federico en Figliozzi (1981) delen de studies in naar hun draagwijdte (micro of macro niveau) en hun aard (theoretisch of operationeel). Op deze manier zijn vier cellen ontstaan. Binnen de cel theoretisch op micro niveau vallen simulatiestudies naar: interacties in sociale situaties, gedrag in kleine groepen, sociometrische groepsstructuur; theoretisch op macro niveau zijn studies naar: verspreiding van informatie binnen een sociaal systeem, kenmerken en structuur van complexe organisaties; binnen de cel operationeel op micro niveau: de mate waarin de operationaliteit van een groep wordt bepaald door persoonlijkheidskenmerken van de leden; operationeel op macro niveau zijn studies naar: te verwachten kiesgedrag, veranderingen in gedrag geïnduceerd door de massamedia of gesprekken, beïnvloedingsprocessen in 'peer groups'.
7. Anderson (1989) legt uit dat het gebruik van frames uit de kunstmatige intelligentie voor sociaal-wetenschappers op twee manieren van groot belang is, namelijk voor de analyse van complexe gebieden en rolstructuren. Een frame-systeem bestaat uit onderling verbonden beschrijvingen van feiten. Een frame beschrijft alle belangrijke eigenschappen van een feit, waaronder waarden, verwijzingen naar een ander frame of instructies voor de verwerking van gegevens.

8. Ook kan simulatie tijdens de onderzoeksopzet worden gebruikt, het betreft dan vooral logistieke zaken: kosten van het onderzoek, vaststellen hoeveel gegevens nodig zijn ten-einde bepaalde conclusies significant te kunnen laten zijn, bepalen van tijd nodig per proefpersoon per meetinstrument.
9. Het Nederlands Historisch Data Archief beoogt gegevens voor historisch onderzoek vast te leggen in computerbestanden. Omdat gegevens op schrift met behulp van scanners vrij gemakkelijk in computers ingevoerd kunnen worden, zijn dergelijke computerbestanden ook zeer wel op te bouwen. Spraakinvoer lijkt voorlopig nog niet mogelijk te zijn.