

PLANNING VAN COMPUTERGEBRUIK IN SOCIAAL WETENSCHAPPELIJK ONDERZOEK.

J.G.A. VINK*

Samenvatting.

De computer neemt, binnen sociaal wetenschappelijk onderzoek, een steeds prominenter plaats in. Deze prominentie heeft voor- en nadelen. Een goede planning van het gebruik van de computer bij het opzetten van het onderzoek kan veel problemen voorkomen. Door gebruik te maken van bijvoorbeeld een systeem als het FSDA-systeem kan men de computerwerkzaamheden vereenvoudigen en versnellen.

Inleiding.

Reeds decennia geleden is er binnen de wereld van het sociaal wetenschappelijk onderzoek een ontwikkeling in gang gezet m.b.t. het gebruik van de computer. In eerste instantie was deze ontwikkeling voornamelijk op het gebied van de kwantitatieve methoden. Voorbeelden hiervan zijn de ontwikkelingen van de standaardpakketten als SPSS en SAS en de programma's voor specifieke technieken, bijvoorbeeld HOMALS en PRINCALS. Voor wat betreft de analyse technieken is er nu ook een ontwikkeling op het gebied van de kwalitatieve methoden met als voorbeelden pakketten voor het analyseren van gegevens uit ongestructureerde interviews, expertsystemen en/of simulatiemodellen.

De laatste jaren is er echter ook een uitbreiding te zien bij andere fasen van het onderzoek. Het mag dan ook vreemd genoemd worden dat veel aandacht wordt besteed aan toepassingen van de computer maar vrijwel nergens aan inpassing van en planning m.b.t. de computer in het onderzoeksproces.

In dit essay wil ik proberen een kleine aanzet te geven om dit hiaat in te vullen. Daarbij zal ik mij vooral concentreren op het gebruik van het systeem van Formele Specificatie voor Data Analyse (FSDA), dat is ontwikkeld op het Instituut voor Maatschappelijke Gezondheidszorg (M.G.Z.) aan de Erasmus Universiteit. Opgemerkt dient te worden dat deze bespreking (nog) niet berust op praktische ervaring met dit systeem maar gericht is op de gedachte die achter een dergelijk systeem ligt.

In de volgende paragraaf zal ik aan de hand van het schema**, dat Segers (1977) van het

* Adres: Sectie Onderzoek Advisering, Faculteit Sociale Wetenschappen, Erasmus Universiteit, Postbus 1738, 3000 DR Rotterdam.

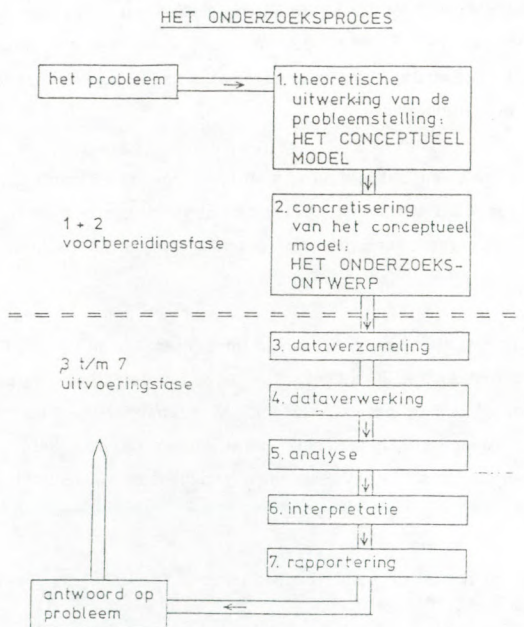
** Bij het gebruiken van het schema ben ik mij ervan bewust dat het een ideaal-typische afbeelding van het werkelijke onderzoeksproces is. Een schematische aanduiding van verschillende fasen, of beter subprocessen, maakt een bespreking echter overzichtelijker.

onderzoeksproces geeft, proberen aan te geven waar in het onderzoek te computer gebruikt wordt. Daarnaast zal ik pogen te verduidelijken dat naast het gebruik van de computer ook aandacht voor het gebruik van de computer in het onderzoeksproces thuis hoort. Vervolgens zal ik mij richten op het FSDA-systeem en de eventuele voor- en nadelen voor de sociaal wetenschappelijke onderzoeker.

Computer in onderzoek.

In het schema van Segers (Figuur 1) worden totaal 7 subprocessen aangegeven. Deze 7 subprocessen worden weer onderverdeeld in 2 'hoofd'subprocessen nl. de voorbereidings- en de uitvoeringsfase. De computer is tot nog toe alleen gezien als een apparaat t.b.v. de subprocessen in de uitvoeringsfase nl. dataverwerking en -analyse (fasen 4 en 5). De laatste jaren heeft hij echter ook toepassingen gekregen in andere onderdelen van het onderzoek.

Figuur 1.



Een voorbeeld van deze toepassingen is de toepassing bij het opstellen van het conceptueel model (fase 1). Door de automatisering van bv. literatuurbestanden zijn de mogelijkheden voor opzoeken van belangrijke literatuur enorm uitgebreid. Was men vroeger afhankelijk van kaartenbakken in een enkele bibliotheek of van verwijzingen uit reeds gelezen literatuur, tegenwoordig kan men gebruik maken van internationale literatuurbestanden waarbij men op diverse criteria verwijzingen kan verkrijgen.

Bij het verzamelen van data (fase 3) wordt tegenwoordig ook gebruik gemaakt van computers. Voorbeelden hiervan zijn het NIPO-Telepanelen de draagbare computers van CBS-interviewers. De invloed op de betrouwbaarheid en de validiteit van de data van deze manier van data verzameling wordt nog onderzocht maar de voorlopige indrukken zijn positief (Nederhof, 1987). Door deze manier van werken wordt de fase van de data invoer bovendien gekoppeld aan de dataverzameling, wat een aanmerkelijke tijdswinst opleverd.

Ook de dataverwerking met computers heeft een gedaanteverwisseling ondergaan. Was het vroeger een kwestie van het eerst coderen en vervolgens maken van bijvoorbeeld ponskaarten en het laten inlezen van de ponskaarten door de computerafdeling, tegenwoordig zijn het coderen en invoeren door gebruik te maken van database management systemen (dbms) tezamen met invoerapplicaties minder separate bezigheden. Behalve de mogelijkheid controle-criteria aan te brengen bij de invoer van geprecodeerde data is het ook mogelijk niet gecodeerde data in te voeren en vervolgens na een eerste analyse met de computer te coderen, of zelfs door de computer automatisch te laten coderen^{*}. Door deze ontwikkeling heeft de onderzoeker meer vrijheid gekregen in het verwerken van zijn data. Die vrijheid bestaat vooral uit de mogelijkheid om vooranalyses en/of tussentijdse analyses van de data uit te voeren.

Voor kwalitatieve analyses met de computer komen er steeds meer mogelijkheden. De ontwikkelingen in de dbms-en m.b.t. de mogelijkheden om met ongestructureerde teksten te werken gaan steeds verder. Daarnaast is het mogelijk om m.b.v. expertsystemen of daarop gelijkende programma's bijvoorbeeld simulatiemodellen toe te passen.

Als laatste ontwikkeling wil ik de tekstverwerking en daaruit voortvloeiend het Desk Top Publishing noemen. De mogelijkheden om zelf m.b.v. de computer publikaties geschikt te maken voor de lezer neemt hand over hand toe. Grafische pakketten en verbeterde afdruk mogelijkheden hebben er bovendien voor gezorgd dat resultaten makkelijk gevisualiseerd kunnen worden. Deze visualisering maakt resultaten inzichtelijker voor grotere populatie. Hierdoor heeft ook in fase 7 de computer naast de gewone tekstverwerking een grotere rol gekregen.

* Bijvoorbeeld het AUTORECODE-commando in SPSS^X v.a. versie 2.2.

De computer in het onderzoeksproces.

Uit de vorige paragraaf blijkt dat de opmars van de computer redelijk sterk is geweest en eigenlijk nog altijd is. Uit de praktijk van het onderzoek blijkt dat men pas gaat nadenken over het gebruik van de computer in het onderzoek wanneer men al in een ver gevorderd stadium van het onderzoek is. In veel gevallen komt de vraag "Hoe ga ik de computer gebruiken?" pas aan de orde wanneer de data al verzameld zijn. Hier wil ik betogen dat men dan veelal te laat is en onnodige problemen oproept m.b.t. verwerking en analyse van de gegevens. Daarnaast heeft men vaak al tijd en geld verspeeld.

Mijn betoog richt zich voornamelijk op het invoeren van een gewoonte om reeds bij de concretisering van het conceptueel model (fase 2) aandacht te besteden aan de vraag of en, indien ja, hoe men de computer wil gaan gebruiken in het onderzoek. Deze centrale vraag kan voornamelijk beantwoord worden door de volgende vragen te beantwoorden:

- Hoe wil men de data gaan verzamelen?
- Welke analysetechnieken wil men gebruiken?
- Hoe organiseert men de bestanden?
- Welke kennis en ervaring heeft de onderzoeker m.b.t. het gebruik van de computer bij data analyse?

De antwoorden op deze vragen staan natuurlijk niet los van elkaar. De keuze van analysetechniek beïnvloedt de eisen die aan data gesteld gaan worden en daardoor indirect de wijze van dataverzameling. De randvoorwaarden waarbinnen de bestandsorganisatie plaatsvindt worden op hun beurt door de eerste twee punten bijgesteld. Het geheel wordt natuurlijk gebaseerd op de probleemstelling en het kennisniveau van de onderzoeker.

Behalve deze onderzoekstheoretische overwegingen zijn er natuurlijk ook praktische overwegingen. Van belang is te weten over welke programma's men kan beschikken en wat de mogelijkheden van die programma's zijn. Welke computer wordt er gebruikt en wat zijn de voor- en nadelen van die computer? Natuurlijk is op dit vlak de financiële factor niet te verontachtzamen. Mankracht en computergebruik kosten geld. De hoeveelheid geld schept randvoorwaarden voor alle keuzen, die hierboven genoemd zijn. De keuze van de hard- en software wordt bijvoorbeeld meestal gereduceerd tot de keuze voor wat er aanwezig is.

Voorkomende problemen.

Onderzoekers houden niet altijd rekening met de in de vorige paragraaf genoemde invloeden. Vaak begint men onderzoek te doen en wanneer men vervolgens bij de fase van de dataverwerking gearriveerd is, gaat men nadenken over de organisatie van de verwerking en analyse

van de gegevens. De computer wordt dan vaak als reddingsboei gebruikt. Deze manier van handelen komt gedeeltelijk voort uit onbekendheid met de mogelijkheden en onmogelijkheden bij het computergebruik. De idee dat door één druk op de knop de resultaten op het bureau liggen is misschien niet expliciet maar bij veel onderzoekers dan toch wel impliciet aanwezig. Men vergeet vaak dat de computer slechts een hulpmiddel is en de onderzoeker de creatieve motor.

Wanneer deze onderzoekers willen gaan werken aan de dataverwerking en -analyse dan blijkt plotseling dat zich problemen voordoen die zij niet voorzien hadden. Eén van de meest voorkomende problemen is de slechte aansluiting tussen de verzamelde data en de technieken die voor de analyse gebruikt worden. Wanneer dit niet van te voren wordt overdacht zit de onderzoeker aan het einde van zijn onderzoek met een probleem dat vaak alleen door kunstgrepen is te verhelpen. Dit komt niet altijd ten goede aan de kwaliteit van de resultaten en de conclusie. De uitvoering kost meestal een extra inspanning die in veel gevallen uit te drukken is in (computer)tijd en derhalve in geld.

Ook bij de keuze van de programma's en de computers blijken problemen de kop op te steken. Wanneer een onderzoeker data heeft verzameld met in zijn achterhoofd de gedachte een bepaald pakket te gaan gebruiken dan kan hij nog vastlopen. Indien hij geen rekening heeft gehouden met de beperkingen van dat pakket, dan is de kans aanwezig dat hij alsnog van programmakeuze moet veranderen. Hierdoor kunnen plotseling veranderingen ontstaan voor de mogelijkheden van analyse.

Als laatste probleem is er de invoer van de data. Wanneer de onderzoeker een dbms wil gebruiken met een aangepaste applicatie voor de invoer van de gegevens dan moet hij met een tweetal dingen rekening houden. Ten eerste kost het maken van een invoerapplicatie tijd. Hoewel het niet de meest tijdrovende programmeeractiviteit is, is het toch aan te raden deze applicatie te ontwikkelen of te laten ontwikkelen op het moment dat de soort en de vorm van de te verzamelen data bekend zijn. Door een parameter-gestuurd ontwerp te maken is aanpassing van de applicatie in een later stadium minder tijdrovend. Een tweede aandachtspunt is de keuze van het pakket waarmee de analyse verricht moet gaan worden. Wanneer dit pakket zeer specifieke eisen stelt aan het formaat van de (meta)gegevens dan zal daar bij het ontwerpen van de invoerapplicatie rekening mee moeten worden gehouden. Bovendien zal men moeten kijken wat de mogelijkheden zijn om de data en beschrijving van de data van het invoerpakket naar het analysepakket te converteren. Levert dit problemen op dan zit er vaak niet veel meer op dan twee maal het zelfde werk te doen, nl. de data specificeren voor beide programma's.

Deze voorbeelden van problemen die kunnen voortvloeien uit een slecht gebruik van de computer in het onderzoek tonen aan dat planning in de tweede fase noodzakelijk is. Tezamen met een kosten-baten analyse kan dit een aanzienlijke besparing van tijd en geld opleveren.

Management informatie systemen.

Voor het verloop van het bovengenoemde proces kan men een parallel trekken naar het ontwerpen van management informatie systemen (M.I.S.). In het kader van een dergelijke parallel wil ik wijzen op de methode die Hartman (1986) Doel-Functie-Data-methodiek noemt. Hij spreekt in het kader van het ontwerpen en invoeren van een M.I.S. over een zevental stappen:

1. Organisatie-analyse.
2. Informatieplan.
3. Automatiseringsplan.
5. Apparatuur- en Programmatuurplan.
6. Activiteitenanalyse.
7. Gegevensanalyse.

Wanneer wij nu voor organisatieanalyse de term ontwikkelen-conceptueel-model in de plaats zetten en voor informatieplan de term operationalisatie dan zal de eerste lijn van de parallel wel duidelijk worden. Hartman wijst vervolgens op het belang van een automatiseringsplan. Dit is belangrijk omdat in een dergelijk plan de keuze wordt gemaakt wat er wel en wat er niet geautomatiseerd moet worden. Vervolgens komt de vraag aan de orde welke hard- en software te gebruiken is in het apparatuur- en programmatuurplan. Voor de praktische uitwerking naar wat er moet gebeuren is er de stap van de activiteitenanalyse en tenslotte in de stap die Hartman gegevensanalyse noemt, komt er een terugkoppeling van de verzamelde informatie over het computergebruik naar de oorspronkelijke probleemstelling en de restant van het onderzoeksontwerp.

Mijns inziens is een dergelijke opvatting over het gebruik van de computer in sociaal wetenschappelijk onderzoek een methode om beter gepland en soepel lopend onderzoek m.b.v. de computer te krijgen. Ik moet daarbij wel aantekenen dat de stappen die Hartman geeft niet zo rigide zijn als hier gebracht en dat ook zeker niet moeten zijn. Zoals reeds boven aangegeven is het goed ontwerpen van een automatiseringsplan voor onderzoek een serie van terug- en vooruitkoppelingen, net zoals Hartman aangeeft voor het ontwikkelen van een M.I.S.

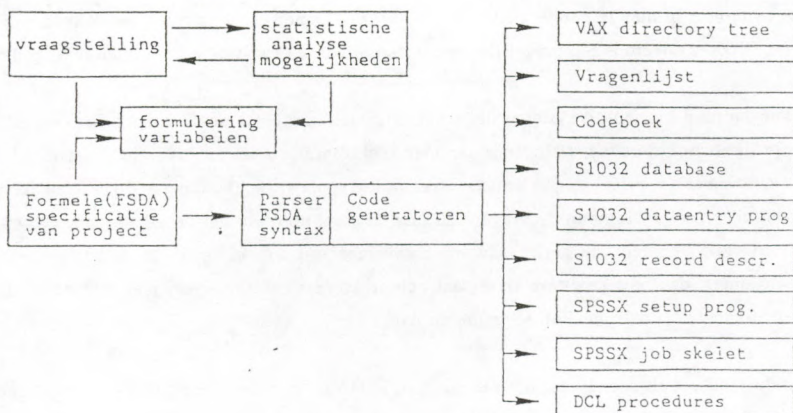
Formele specificatie voor data analyse.

Een onderdeel van een beter gebruik van de computer in een onderzoek kan zijn het formeel specificeren van de variabelen die in dat onderzoek gebruikt gaan worden. Bij M.G.Z. is een systeem ontwikkeld dat op dit idee is gebaseerd. In figuur 2 staat een schematische afbeelding van dit FSDA-systeem (Segaar, 1987).

In dit essay zal ik mij voornamelijk richten op het linker gedeelte van het schema (Figuur 2). In dit gedeelte wordt de gedachte achter het systeem schematisch weergegeven. Het rechter gedeelte is de uitwerking van het systeem m.b.v. de op de computer van de EUR aanwezige software. Naast de in dit schema genoemde mogelijkheden werkt men nog aan uitbreidingen.

De eerste koppeling die van belang is is die tussen de vraagstelling en de mogelijke analyse technieken om deze te beantwoorden. Bij deze koppeling komt de vraag aan de orde of men wel gebruik zou moeten maken van een computer of niet. In het schema worden statistische analyse technieken genoemd. Deze verwijzing komt voort uit de oorsprong van het systeem. Het is namelijk ontwikkeld ter versoepeling van een kwantitatief gericht onderzoek. De eerste koppeling is echter ook van belang voor kwalitatief onderzoek.

Figuur 2.



Een volgend (dubbel)verband dat wordt gelegd is dat tussen de technieken en vraagstelling enerzijds en de variabelen anderzijds. Wanneer deze verbanden zijn uitgewerkt kan men in principe een keuze maken voor de wijze van gegevensverzameling. Deze keuze moet vervolgens

teruggekoppeld worden op de drie eerdere keuzen. Immers de vorm van de data verzameling beïnvloedt de vorm en de inhoud (van de formulering) van de variabelen.

Wanneer de eerste cirkel, de operationalisatie, rond is komt in het schema de volgende stap: de formele specificatie. In deze stap worden de variabelen gedefiniëerd op basis van inhoud en vorm. In tegenstelling tot de meeste programma's worden zij bij het FSDA-systeem niet gedefiniëerd op basis van de vorm waarin zij worden opgeslagen door de computer, bijvoorbeeld text of integer, maar op basis van hun niveau, nominaal, ordinaal of interval. Op deze manier moet de onderzoeker reeds bij het begin van zijn onderzoek nadenken over de (on)mogelijkheden voor gebruik in een analyse.

Na deze eerste fase komt er een koppeling naar de uitvoering met behulp van het FSDA-systeem. M.b.v. dit systeem worden zoals in het rechter gedeelte van het schema is te zien een aantal zaken gegenereerd. De bestandsorganisatie, vragenlijsten, codeboeken, etc. worden automatisch op basis van de specificatie gegenereerd. Hierdoor zijn alle delen van het onderzoek waarbij de computer gebruikt wordt op elkaar aangesloten.

Conclusie.

Wanneer reeds bij de fase van het onderzoeksontwerp nagedacht wordt over het gebruik van de computer binnen dat onderzoek kunnen veel problemen en extra werk voorkomen worden. Het FSDA-systeem is een mogelijke eerste stap om deze gewoonte in te voeren in de praktijk.

Doordat men binnen dit systeem slechts eenmaal de variabelen hoeft te specificeren voor het gebruik in meerdere toepassingen spaart men veel computerwerk en -tijd. Het systeem verzorgt bovendien een overzichtelijke bestandsorganisatie en de ontwikkeling van de vragenlijst en de eventuele verwerking van de vragenlijst met de computer worden op elkaar aangesloten. De specificatie op basis van het niveau van de variabelen i.p.v. op basis van de opslagvorm kan voorkomen dat onderzoekers er te laat achter komen dat een specifieke techniek niet te gebruiken is op de door hen verzamelde data.

Afsluitend concludeer ik dat het gebruik van FSDA of in de invoering van de achterliggende gedachte in de onderzoekspraktijk het verloop van onderzoek kunnen versnellen. Als bijkomend voordeel kunnen de kosten van computergebruik door efficiënter gebruik van ruimte en tijd aanzienlijk beperkt worden.

Literatuur.

- Hartman, W., Het ontwerpen van informatiesystemen, Deventer, 1986
 Nederhof, A.J., De grote onbekende, in: Sociologische Gids, 1987/2

- Segaar, R., Het FSDA systeem: Formele Specificatie voor Data Analyse, in
 Convex courier, jaargang 3, nummer 6, Rotterdam/Tilburg/Maastricht/Wageningen, 1987.
- Segers, J.H.G., Sociologische onderzoeksmethoden, Assen/Amsterdam, 1977

Ontvangen: 16-05-1988

Geaccepteerd: 03-07-1989