

Cybersociologie

Een beknopt kookboekje

Voorwoord

Het internet wordt inmiddels gebruikt door mensen uit alle bevolkingsgroepen. Ook het world wide web is doodgewoon geworden. De tijd is dan ook voorbij dat het internet, en sinds 1995 ook het world wide web, model stonden voor een nieuwe, fundamenteel vrije vorm van communicatie, radicale democratisering, en de realisering van de droom van het wereldbrein. Dat maakt het verschijnsel internet voor sociale wetenschappers natuurlijk pas echt interessant. Dit kookboekje beoogt hulp te verschaffen aan de sociale wetenschapper die gebruik wil maken van het internet en het web in het onderzoek. Het is geen uitputtend overzicht van alle literatuur op dit gebied, maar kan worden gebruikt als introductie voor wie het internet nog niet vaak voor wetenschappelijk onderzoek heeft benut. De belangrijkste problemen en hinderpalen komen aan de orde en er worden enkele voorbeelden van succesvolle recepten gegeven. Het kookboekje besluit met een literatuurlijst voor de onderzoeker die zich verder wil bekwalen in de sociaal-wetenschappelijke kookkunst met internet.

Ingrediënten

Het internet: het publiek toegankelijke wereldwijde netwerk van computernetwerken (Huizer 2002). Hieronder verstaan we in dit kookboekje ook het gedeelte waarvoor moet worden betaald of dat met een wachtwoord wordt beschermd. De afgeschermd intranetten van bedrijven of instellingen vallen

hier echter niet onder. De commerciële of militaire computernetwerken zullen we eveneens buiten beschouwing laten.

Het world wide web: de dominante interface van gebruikers met informatie op het internet (Berners-Lee 1999). Hoewel het www, gebaseerd op het http-protocol slechts één van de mogelijke internet-protocollen is, is het de dominante vorm van toegang tot het internet geworden. Andere protocollen (zoals ftp, usenet nieuwsgroepen, online educatiesystemen, en kletskanalen zoals Yahoo Chat www.yahoo.com) kunnen nu ook via het www worden benaderd. Zozeer zelfs dat het internet en het world wide web in het dagelijks spraakgebruik vrijwel synoniem zijn geworden, hoewel dit strikt genomen onjuist is.

Methodologie: de wijze waarop sociaal-wetenschappelijk onderzoek in de praktijk wordt uitgevoerd. In dit kookboekje hanteren we een iets breder begrip dan in veel literatuur over methodologie gebruikelijk. Onder methodologie verstaan we niet alleen het geheel van voorschriften, normen, methoden en technieken dat garant geacht wordt te staan voor valide, wetenschappelijk interessante, en bruikbare sociaal-wetenschappelijke kennis. We betrekken er uitdrukkelijk de wijze bij waarop sociale wetenschappers in de praktijk gebruikmaken van het internet in hun onderzoeksmethode, ook waar dit indruist, of lijkt in te druisen, tegen de gangbare normatieve methodologische voorschriften. De grens tussen methodologie en theorie is overigens diffuus. In ons bestek zijn ook meta-theoretische posities ten opzichte van het internet (o.a. 'revolutie' (Castells 1996) versus 'evolutie' (Webster 1995)) interessant.

De voorbereiding

De internethype is verdwenen (met dank aan de aandelenbeurzen) en ook de opgeklopte verwachtingen over de methodologische potentie van het internet

1 Een interessant overzicht van informatie over oorlogsvoering wordt gegeven in Davies 2002.

2 Een introductie tot file transfer protocol is te vinden op http://searchnet.working.techtarget.com/sDefinition/0,,sid7_gci213976,00.html.

3 Een introductie tot usenet is op het web te vinden op <http://www.ibiblio.org/usenet-i/usenet-help.html>.

4 Er zijn verschillende online-systemen die door universiteiten worden gebruikt. De Amsterdamse universiteit gebruikt blackboard <http://blackboard.ic.uva.nl/>.

en het web worden nauwelijks meer serieus genomen. Ondanks de belofte van het overbodig maken van de wetenschappelijke uitgevers, is de wijze van wetenschappelijk publiceren vooralsnog niet omver geworpen door e-tijdschriften (Taubes 1996) of web-portals, gestructureerde websites die toegang bieden tot het domein van specifieke informatie.⁵ Ook de droom van een radicaal nieuw type vertoog met de komst van hypertext is niet of nauwelijks gerealiseerd. Hypertext werd ver voor de opkomst van het web door Ted Nelson geïntroduceerd in zijn *Literary Machines* (Nelson 1981) als een nieuwe, vorm van schrijven en publiceren. Het bestaat uit tekstblokken die zijn verbonden door elektronische links. HTML⁶, de meest gebruikte opmaaktaal voor webpagina's, is een voorbeeld van hypertext. Volgens Nelson biedt hypertext een grotere vrijheid aan de lezer, die de links immers naar eigen keuze kan volgen. In dezelfde trant is hypertext door een aantal auteurs gevierd als een nieuwe vorm van intertekstualiteit door het samenkomen van postmoderne theorie en computertechnologie (Landow 1992). Niettemin heeft hypertext het lineaire argument niet vervangen.

Kennis produceren is in verreweg de meeste vakgebieden bovendien nog steeds een langzaam proces. Het internet heeft niet alles sneller gemaakt. Ook heeft het geen walhalla geschapen waarin een overvloed aan bruikbare data voor elke onderzoeker vanaf de desktop bereikbaar is, die nog slechts hoeven te worden geanalyseerd. Er zijn weliswaar veel gegevens beschikbaar, maar zoals we later in dit kookboekje zullen zien, zijn deze doorgaans niet geschikt voor onmiddellijk gebruik in sociaal-wetenschappelijk onderzoek. De 'overvloed van informatie' waarvoor al sinds de negentiende eeuw met de regelmaat van de klok wordt gewaarschuwd, heeft de noodzaak van het zorgvuldig opbouwen van dataverzamelingen niet overbodig gemaakt. In 2000 bevatte het world wide web naar schatting ongeveer 21 terabyte statische informatie (webpagina's) (Lyman & Varian 2000). Dit deel van het web verdubbelt elk jaar. Het 'onzichtbare web' is vermoedelijk vele malen groter. Dat is dat deel van het web dat op basis van de huidige technologie niet met zoekmachines doorzoekbaar

5 Een voorbeeld van een portal voor cybersociologen is de Sociosite aan de UvA: <http://www.pscw.uva.nl/sociosite/index.html>. Webuse is een geavanceerde portal voor internetonderzoekers aan de University of Maryland waarin het accent wat meer op kwantitatief onderzoek en datasets ligt: <http://www.webuse.umd.edu/> (Neustadt et al. 2002). Een voornamelijk mislukte poging een portal op te zetten over de rol van de Nederlandse overheid in de informatiesamenleving is te vinden op de website van Infodrome: <http://www.infodrome.nl/>.

6 <http://www.w3.org/MarkUp/Guide/>.

is, zoals dynamisch gegenereerde webpagina's, databases, beeldbestanden en dergelijke (Sherman & Price 2001). E-mail bevat naar schatting vijfhonderd maal zoveel informatie als de statische webpagina's. De schattingen variëren maar per jaar schijnen op dit moment 610 miljard e-mails te worden verzonden (Lyman & Varian 2000). De meeste bits op het internet zitten overigens in videobeelden. Met andere woorden, verreweg de meeste data op het internet zijn niet zonder meer geschikt om in sociaal-wetenschappelijke dataverzamelingen te worden opgenomen.

Dit betekent niet dat de onderzoeker internet en het web ongestraft kan negeren. Daarvoor is het een te opvallend en interessant maatschappelijk verschijnsel. Het sociaal-wetenschappelijk onderzoek naar de gevolgen van de opkomst van internet heeft zich de afgelopen jaren snel ontwikkeld (dit nummer is daar onder meer getuige van) (Wellman & Haythorntwaite 2002). Bovendien biedt het medium ook interessante methodologische mogelijkheden. Wel doet de internetonderzoeker er verstandig aan het gebruik van het internet zorgvuldig en zo precies mogelijk in de onderzoeksopzet op te nemen. Houd daarbij rekening met het complexe karakter van het internet. In feite is het misleidend om van het internet of het web als eenduidig object te spreken (Wakeford 2000). Het gaat eerder om een veelheid aan communicatiemogelijkheden, vormen van amusement, en een grote variëteit aan typen informatie, die elk op zichzelf specifieke karaktertrekken hebben. Er is dan ook geen eenduidig antwoord op de vraag hoe het internet als onderzoeksinstrument kan worden benut. Wel is de vraag relevant hoe de bestaande sociaal-wetenschappelijke onderzoeksmethoden gebruik kunnen maken van aspecten van het internet. Dat kan uiteindelijk leiden tot het ontstaan van nieuwe onderzoeksmethoden, ook al is het nog wat vroeg om daar zekerheid over te hebben.

Juist op basis van een sceptische houding komen de interessante mogelijkheden tot daadwerkelijke vernieuwing van sociaal-wetenschappelijke methoden en technieken in beeld. Daarbij is het nuttig om in het gebruik van 'het internet' als element in het onderzoeksinstrumentarium onderscheid te maken tussen verschillende fasen van het onderzoek:

- literatuur- en informatie verzamelen;
- indicatorconstructie (bij kwantitatief onderzoek) en datacollectie;
- data-analyse en -interpretatie;
- presentatie en publicatie van de onderzoeksresultaten.

Informatie verzamelen

Het opzoeken van informatie op het world wide web is zo gemeengoed geworden dat we daar kort over kunnen zijn. Door de ontwikkeling van de digitale bibliotheek⁷ (Harter 1997; Carpenter, Shaw & Prescott 1998; Fox & Urs 2002), de digitalisering van de wetenschappelijke uitgeverij en wetenschappelijke verenigingen is alle wetenschappelijke literatuur via het web beschikbaar, althans voor wie een abonnement heeft. Dat spaart een loopje naar de bibliotheek uit. In toenemende mate worden ook de gratis elektronische tijdschriften in de bibliotheekcollectie opgenomen.⁸ De meeste gevestigde tijdschriften hebben een hybride karakter: ze zijn zowel elektronisch als gedrukt verkrijgbaar. Er is geen sprake van een uniforme trend naar e-tijdschriften (Kling & McKim 2000). Het internet heeft met andere woorden vooral meer interfaces tot de bestaande literatuur geschapen. Dat is niet altijd gepaard gegaan met een grotere toegang tot die literatuur.

Op het gebied van de 'grijze literatuur' heeft de opkomst van het web meer betekend. Proceedings, onderzoeksrapporten en lopende onderzoeksprojecten waren altijd moeilijk te vinden. Nu heeft menige onderzoeksgroep een website waar deze informatie vrij verkrijgbaar is. We weten overigens nog niet in welke mate dit een algemene trend is en op welke punten onderzoekers uit verschillende vakgebieden een verschillend webgedrag vertonen (Voorbij 1999). Toch is het de moeite waard om enkele uren (of dagen: zoeken op het web kost vrijwel altijd meer tijd dan de onderzoeker verwacht) te besteden aan het opzoeken van actuele projecten op het web. De websites bevatten ook contactinformatie en zijn daarmee een hulpmiddel in het vinden van potentiële partners voor samenwerking. Omdat samenwerking ook in de sociale en geesteswetenschappen in toenemende mate van belang wordt (Bijker & Peperkamp 2002; Katz & Martin 1997; Newman 2000; Olson, Finholt & Teasley 2000; Sudweeks & Allbritton 1996; Walsh & Maloney), kan deze zoekfase belangrijke resultaten opleveren. Vooral als het een nieuw onderzoeksdomein betreft waarvan de relevante groepen nog niet in kaart zijn gebracht.

7 Het NSF-programma dat een belangrijke rol heeft gespeeld in de ontwikkeling van het idee van een digitale bibliotheek is te vinden op <http://www.ehr.nsf.gov/ehr/duel/programs/nsdl/>.

8 De Koninklijke Bibliotheek in Den Haag beheert een digitale collectie van online tijdschriften: http://www.kb.nl/kb/resources/frameset_collecties.html.

Datacollectie en indicatorconstructie

In deze sectie gaan we achtereenvolgens in op:

- het vinden van kant en klare data;
- de aard van data over het internetgebruik;
- het verzamelen van data over internetgebruik;
- het benutten van internetcommunicatie voor het aanleggen van een dataverzameling;
- internet en indicatorconstructie.

Het vinden van kant en klare data

De onderzoeker dient zich in de eerste plaats te realiseren dat de meeste informatie niet op het publieke web is te vinden, maar in andere delen van het internet. Het zogeheten onzichtbare web (Sherman & Price 2001) bestaat zoals gezegd vooral uit dynamisch gegenereerde webpagina's, databanken en de institutionele intranetten. Overweeg in welke gespecialiseerde databanken relevante data te vinden kunnen zijn. Overzichten van deze databases zijn gemakkelijk via het internet te raadplegen. De meesten hebben een web-interface ontwikkeld. Door overheden gefinancierde databanken zijn vaak vrij toegankelijk.⁹ Wel wordt het gebruik van data over menselijk gedrag uiteraard beperkt door onder andere de privacywetgeving. In Nederland moeten de microdata van het CBS daarom worden opgevraagd via het Wetenschappelijk Statistisch Agentschap.¹⁰ Informatie over dit soort regelingen is doorgaans beschikbaar op het web. Ten slotte is een aantal data-archieven opgericht om hergebruik van data in de sociale wetenschappen te bevorderen,¹¹ ook zij zijn in toenemende mate via het web benaderbaar. Veel sociaal-wetenschappelijke data bevinden zich echter in de 'schoenendozen' van individuele onderzoekers of onderzoeksgroepen. Het delen van data blijkt vooralsnog een problematisch proces, gehinderd door competitie tussen onderzoekers, gebrek aan standaarden voor dataformatering en opslag, en niet te vergeten gebrekkige financiering van geavanceerde informatiediensten (OECD 1998; Wouters & Schröder 2000;

9 Dit geldt bijvoorbeeld voor het US Patent Office <http://www.uspto.gov/>, het Europese Patentbureau <http://www.european-patent-office.org/>, Europese statistieken via Eurostat <http://europa.eu.int/comm/eurostat/>.

10 http://wsa.magw.nl/index_nl.htm.

11 Een goede voorbeeld hiervan is het Britse Data Archive <http://www.data-archive.ac.uk/>. In Nederland kennen we het Steinmetz archief http://www.niwi.knaw.nl/nl/dd_star/dd_star.htm.

Schroeder 2001; Wouters 2002; Koslow 2002; Wouters & Schröder 2003; Wouters & Beaulieu 2002).

Het hierboven genoemde benutten van internetcommunicatie om partners voor samenwerking te vinden, kan uiteraard wel zeer nuttig zijn voor het verkrijgen van toegang tot data.

Data over het internetgebruik

In de tweede plaats komt de vraag aan de orde of het internet danwel het web zelf relevante data kunnen opleveren. Hiermee doelen we op de data die worden gegenereerd door het gebruik van het internet als zodanig.¹² De onderzoeker realiseert zich dat internetgebruikers doorgaans niet representatief zijn voor de doelgroep. De regelmatig verschijnende NUA Internet surveys¹³ geven een beeld van het internetgebruik per land op basis van technische en sociaal-economische gegevens. Dit soort rapporten kan worden benut om een inschatting te maken van het gedeelte van het voor het onderzoek relevante publiek dat daadwerkelijk een of meer internetdiensten benut. De al eerder genoemde webportal voor internetonderzoekers webuse¹⁴ bevat ook een schat aan relevante survey data met betrekking tot de sociale rol van het internet en het web. De onderzoeker dient zich te hoeden voor de verleiding van de gemakkelijk beschikbare virtuele data. Sociologisch en etnografisch onderzoek heeft voldoende duidelijk gemaakt dat internetgebruik altijd is ingebed in het analoge, off line leven van alledag (Woolgar 2002; Wyatt, Henwood, Miller & Senker 2000; Dutton 1996; Mann & Stewart 2000; Miller & Slater 2000). Virtuele data hebben echter de neiging belangrijke aspecten van deze context onzichtbaar te maken. Zo verduistert de exponentiële stijging van internetgebruik in een aantal maatschappelijke sectoren de mate van niet-gebruik van het internet, een aspect dat echter onmisbaar is voor de interpretatie van de maatschappelijke rol van het internet (Wyatt, Thomas & Terranova 2002). De onderzoeker moet zich dan ook steeds afvragen in welke context de data zijn gegenereerd, naar welke activiteit of proces zij kunnen verwijzen, en welke aspecten van het te onderzoeken fenomeen onzichtbaar zijn gemaakt. Hoe belangrijk dit laatste is, hangt uiteraard af van de onderzoeksvraag.

12 De literatuur over dit onderwerp wordt besproken in Molyneux & Williams 1999.

13 <http://www.nua.ie/surveys/>.

14 <http://www.webuse.umd.edu/>.

Het is onbetwistbaar dat het internetgebruik zelf een interessant nieuw maatschappelijk verschijnsel is. Het maakt een nieuwe combinatie van eigenschappen van oudere massamedia mogelijk (Thompson 1995; Slevin 2000). Naarmate de onderzoeksvraag zich meer beperkt tot het domein van het internetgebruik zelf, zal de aard van de internetdata voor het onderzoek geschikter zijn. De reikwijdte van de conclusies is dan uiteraard navenant beperkt. Naarmate het onderzoek zich echter breder opstelt, dienen de internetdata doorgaans aangevuld te worden met andere gegevens. Ook het aggregatieniveau van de data is van belang. Kwantitatieve data op een hoog aggregatieniveau kunnen vaak pas in combinatie met microdata goed worden geïnterpreteerd.

De onzes inziens meest verbreide nieuwe verschijnselen die internetgebruik met zich mee brengt, het onderzoeken waard, zijn:

- Internetcommunicatie, waaronder elektronische post (e-mail) en kletskanalen (chat channels en chat boxes). E-mail is een vorm van communicatie die een mix lijkt van het telefoongesprek en de vrijwel uitgestorven brief. E-mail is uitgebreid onderzocht en als gevolg daarvan weten we vrij veel over e-mailgebruik. Het is de meest gebruikte internettoepassing. Door het vluchtige en privé-karakter is het verkrijgen van de e-maildata zelf niet eenvoudig. Doorgaans vereist dit toestemming van de betrokkenen. Een nieuwe vorm van communicatie op basis van e-mail is de e-maillijst waarop individuen zich kunnen abonneren, doorgaans gratis. De e-maillijst levert – in contrast tot het synchrone karakter van kletskanalen – asynchrone data op met betrekking tot groepscommunicatie. Naarmate de e-maillijst een explicieter publiek karakter heeft, is het verzenden van e-mail aan de lijst meer vergelijkbaar met publiceren. Analyse van deze e-mails wordt navenant minder bemoeilijkt door overwegingen van privacy-bescherming. De onderzoeker kan er echter niet van uitgaan dat alle e-maillijsten zo publiek zijn. Kletskanalen zijn ook vluchtig. Onderzoekers kunnen dit type data wel eenvoudig opslaan door in te loggen en simpelweg alle tekst die op het scherm versijnt te downloaden. Er is inmiddels ook software ontwikkeld die zo'n 'screendump' vereenvoudigt. Ook hierbij geldt dat niet zonder meer kan worden uitgegaan van een publiek karakter van deze data.
- Persoonlijke expressie en exhibitionisme op het web – zoals de persoonlijke website, de webcam en het webdagboek (*blog*). Websites bestaan sinds 1995,

de blog is het nieuwste verschijnsel (Mortensen & Walker 2002).¹⁵ Dit type expressie op het internet en het web is een fascinerend, complex geheel waarvan het systematisch bestuderen nog in de kinderschoenen staat. Het verzamelen van de data is technisch niet zo eenvoudig door de aard van de bestaande zoekmachines (zie hieronder). Het is raadzaam hier voldoende tijd voor te nemen. Het aanleggen van de dataverzameling is in principe een cyclisch proces waarin meerdere bronnen worden gecombineerd. Besteed veel aandacht aan het corrigeren van fouten. De wijze waarop dit het best kan gebeuren hangt uiteraard af van de vraagstelling van het onderzoek. Meestal zijn de bestaande methodologische handboeken een goed uitgangspunt. In kwantitatief onderzoek kan het bestaan van meerdere versies en van mirror-sites roet in het eten gooien. Voor een kwalitatieve analyse van expressie en exhibitionisme is het cruciaal de context waarin de website wordt geproduceerd te kennen. Veel exhibitionisme is commercieel geïnspireerd, toch kan daar niet bij voorbaat van worden uitgegaan. Wie conclusies wil trekken over de betekenis van dit type websites voor de gebruiker ontkomt doorgaans niet aan het houden van interviews.

- Het gebruik van internet voor informatieverzamelen. Vooral over ziekte en gezondheid wordt het internet massaal geraadpleegd. Traditioneel gebruikersonderzoek kan hiertoe worden gecombineerd met de analyse van logfiles, hetzij aan de kant van de informatieaanbieder (server) hetzij door logsoftware in te bouwen in de computer van de gebruiker (cliënt). Bibliotheek- en informatiewetenschappers hebben de technische kant van dit proces al aardig in kaart gebracht.¹⁶ Informatie verzamelen is echter niet een louter rationeel proces. De emotionele aspecten en de directe gevolgen van dit gedrag hebben tot nog toe voorzover we weten minder aandacht gekregen.
- Het gebruik van internet in het opbouwen van sociale netwerken, van nieuwsgroepen tot virtuele gemeenschappen. De aandacht voor de virtuele gemeenschap is van het begin af aan groot geweest (Steinmüller 2002; Rheingold 1994; Davenport & Hall 2002). Na een korte periode waarin

15 Een interessante portal over Blogs is gepubliceerd door Alex Halavais, verbonden aan de School of Informatics, State University of New York at Buffalo: <http://alex.halavais.net/news/>.

16 De jaarlijks uitkomende Annual Review of Information Science and Technology, uitgebracht door de American Society for Information Science and Technology geeft doorgaans een goed beeld. Het tijdschrift JASIST publiceert op hetzelfde gebied.

hooggespannen verwachtingen over de bevrijdende werking van de virtualiteit leken te overheersen, wordt de onderzoeksagenda alweer geruime tijd overheerst door precies, genuanceerd sociologisch en mediaonderzoek (Garton, Haythornthwaite & Wellman 1997; Jankowski 2001; Van den Beselaar, Tanabe & Ishida 2002). Methodologisch staat dit doorgaans in de traditie van de sociale netwerkanalyse, het mediaonderzoek of van de etnografie. Het internetgebruik maakt het mogelijk een deel van de netwerkdata met geautomatiseerde processen te verzamelen. Dit vereist toestemming van de betrokkenen. Het gaat daarbij meestal om logdata, hyperlinkdata, de inhoud van websites, en e-mailverkeer.

- Groepsspecifieke vormen van internetgebruik. Hieronder verstaan we het gebruik van internet als publicatiemedium door professionele uitgevers en de wijze waarop specifieke beroepsgroepen internetinstrumenten in hun werkwijze toepassen. Vrijwel altijd bouwt dit voort op al bestaande routines. Zo was de befaamde preprint server (<http://xxx.lanl.gov>) een vertaling van het met tante Pos versturen van manuscripten (Paul Ginsparg 1996). Radio-astronomen bouwen een snel netwerk zodat ze hun data niet meer per vrachtwagen hoeven op te sturen. Dit schept nieuwe mogelijkheden voor de constructie van virtuele onderzoeksinstrumenten. Methodologisch vereist het onderzoek naar dit groepsspecifieke gebruik het verzamelen van zowel on line- als off linedata.
- Het gebruik van internet als amusementsmedium. De bulk aan data op het internet wordt gegenereerd ten behoeve van amusement (Lyman & Varian 2000) in de vorm van films, video en computerspelletjes. Dit is gerelateerd aan het gebruik voor expressie en exhibitionisme. Ook het onderzoek kent vergelijkbare methodologische uitdagingen als hierboven genoemd. De ontwikkeling van het internet zelf wordt de komende jaren naar verwachting niet meer gedreven door militaire of academische doeleinden maar door de amusementsindustrie.
- Commerciële activiteiten via het internet, zoals veilingen, in- en verkoop van goederen, business to business-relaties en geldelijke transacties. Veel van de interessantste activiteiten op dit gebied gebeurt door middel van beschermde netwerken (zoals de bancaire netwerken) en valt buiten het bestek van dit kookboekje. In het onderzoek naar e-commerce staat de vraag vaak centraal hoe vertrouwensrelaties on line kunnen worden opgebouwd en gehandhaafd (Lunn & Suman 2002).

Alvorens in te gaan op het probleem van het verzamelen van data over internetgebruik, willen we opmerken dat de befaamde tweedeling tussen kwantita-

tieve en kwalitatieve data met betrekking tot internet wellicht niet zo scherp is (Sudweeks & Simoff 1998). Het aardige van internetdata is dat teksten en getallen gemakkelijk samengaan. Het is relatief eenvoudig grote hoeveelheden teksten te verzamelen en die vervolgens getalsmatig te karakteriseren. Onderzoek waarin de nadruk op kwantitatieve analyse ligt, kan anderzijds snel voorbeelden van bepaalde verschijnselen in een tekst illustreren. Dit raakt echter al het domein van de data-analyse. Eerst dienen we in dit kookboekje in te gaan op enige specifieke problemen met het verzamelen van internetdata.

Zoekmachines en robots

Voor veel van bovengenoemd onderzoek zijn data over het gebruik van het world wide web relevant. Het bijeenbrengen van deze data kampt met twee typen problemen:

- Problemen veroorzaakt door de structuur van het web;
- Problemen veroorzaakt door de software nodig om de data te verzamelen.

De structuur van het web

Het web is een netwerk van websites, gedeeltelijk met elkaar verbonden, waarvan de structuur omstreden is. Een stroming in de informatiewetenschap heeft geconcludeerd dat het web een small world-netwerk is (Watts 1999). Daar staat een groep tegenover die een bow-tie-model heeft voorgesteld¹⁷, waarin een deel van het netwerk geïsoleerd is. In een small world-netwerk kan elke knoop uiteindelijk vanuit elke andere worden bereikt. Dat geldt echter niet voor het concurrerende model. Het zal duidelijk zijn dat dit grote gevolgen heeft voor de strategie van data verzamelen. Cruciale eigenschap van het web is dat er wel sprake is van standaardisatie van communicatie- en dataprotocolen, maar er is geen centrale regie. Het web groeit van onderop doordat websites worden gepubliceerd en verdwijnen. De complexiteit van het netwerk leidt op het technische vlak tot zelforganisatie (Adamic & Huberman 2000; Huberman & Adamic 1999). Op individueel niveau betekent dit dat het fundamenteel onzeker is of een bepaalde connectie in het netwerk op zeker moment kan worden benut. Dit leidt ertoe dat naar schatting ongeveer eenderde van de websites instabiel is (Koehler 1999). Er is een groot verloop, na zes jaar is nog maar een minderheid van de websites beschikbaar (Bar-Ilan 2001; Lawrence et al. 2001). Deze gegevens gelden voor websites in het algemeen en zijn met grote onzekerheid omgeven. Het is vooralsnog niet bekend of specifieke ver-

17 <http://www.comp.nus.edu.sg/~liub/>.

zamelingen websites stabiel zijn. Een probleem is bovendien dat de grootte van het web onbekend is, dit geldt ook voor het internet als geheel. Uiteraard is wel bekend welke technische infrastructuur het internet en het web draagt, maar aantallen webpagina's en de hoeveelheid informatie blijken moeilijk te schatten (Lawrence & Giles 1999a; Lyman & Varian 2000).

Deze karakteristieken van het web leiden tot de volgende problemen voor sociaal-wetenschappelijke dataverzameling:

- De totale populatie is onbekend. Dit maakt het niet goed mogelijk op basis van louter internetdata een steekproef te trekken. Wel is het goed denkbaar om steekproeven te nemen op basis van een combinatie met bestaande, van internet onafhankelijke, gegevens.
- Websites zijn instabiel en de mate van instabiliteit is op het niveau van de website onvoorspelbaar. Het kan daarom raadzaam zijn gwebsites in hun geheel op te slaan en zorgvuldig te annoteren. Dit kan wel tot ethische problemen leiden. Het maakt wel verschil of een persoonlijke website regelmatig wordt bekeken of systematisch wordt opgeslagen voor analyse. In het laatste geval dient het vragen van toestemming serieus te worden overwogen.
- De wijze waarop het zoekproces begint kan grote invloed uitoefenen op het eindresultaat. Daarom kan het raadzaam zijn meerdere ingangspunten te kiezen. Ook dan blijft het echter onzeker of alle relevante delen van het web zijn doorzocht.
- Er zijn talloze dubbelgangers op het web. Hierop dient de dataverzameling te worden gecorrigeerd. Mirror-sites hebben een andere rol dan de originele website, en het is niet altijd helder welke website het origineel is. Van documenten slingeren vaak verschillende versies op het web rond. Het web wordt nu eenmaal niet opgeruimd. De onderzoeker doet er daarom verstandig aan voldoende tijd in te ruimen voor het opschonen van het resulterende databestand.
- De kwaliteit van de data kan niet eenduidig uit de on line context worden bepaald. Kwaliteit is uiteraard een relatief begrip zowel met betrekking tot het object als de vraag van het onderzoek.
- Hiermee hangt samen dat de off line context van de internetdata niet eenduidig uit de on linedata kan worden bepaald. Wel geven on linedata vaak meer indicaties dan de gebruikers zich realiseren (*hostname*, patroon van activiteit). Vaak zijn deze indicaties overigens nuttiger voor marktonderzoek dan voor sociaal-wetenschappelijk werk.
- Toegang tot digitale data wordt altijd gemedieerd door software. In langetermijnonderzoek dient de onderzoeker daarom niet alleen de ruwe data

maar ook de software nodig voor het gebruik van die data op te slaan en te annoteren. Denk aan verschillende versies van browser software, tekstverwerkers, web- en desktop publishing-software.

Softwareproblemen

De meeste websurfers gebruiken zoekmachines om informatie op te sporen. Zij kunnen ook worden gebruikt om dataverzamelingen aan te leggen. Belangrijkste probleem daarbij is dat de betreffende algoritmes bedrijfsgeheim zijn. De gebruiker van een zoekmachine zoekt niet op het web maar in de databank van de zoekmachine. De onderzoeker weet niet welke criteria zijn gebruikt in de opbouw van die databank (door de *crawler*) en evenmin is transparant op welke wijze de informatie wordt gepresenteerd en geordend. Informatie- en bibliotheekwetenschappers hebben uitvoerig onderzoek verricht naar de prestaties van zoekmachines. In dit bestek zijn vooral de volgende problemen relevant:

- Geen enkele zoekmachine dekt een substantieel deel van het web (Lawrence & Giles 1999b). Het is raadzaam in ieder geval een combinatie van meerdere zoekmachines te gebruiken (www.searchenginewatch.com).
- Zoekmachines hebben specifieke voor- en afkeuren. Zij kunnen bepaalde websites systematisch negeren zonder dat de gebruiker dat eenvoudig kan zien (Introna & Nissenbaum 2000).
- Zoekmachines veranderen hun algoritmes regelmatig, voortgedreven door de concurrentie en webpolitieke overwegingen (Marres & Rogers 2000; Introna & Nissenbaum 2000).
- De zoektaal (mogelijkheid Boolean operators, aangeven prioriteiten) varieert per zoekmachine. Een string dient dan ook steeds te worden aangepast aan de betreffende zoekmachine. Niet elke zoekmachine heeft dezelfde mogelijkheden.
- De criteria voor de ordening van de data zijn niet transparant. Als gevolg van het opvragen van een andere ordening van de resultaten kan zelfs het totaal aantal hits bij dezelfde zoekstring veranderen.
- De zoekresultaten zijn niet altijd stabiel, ze kunnen zelfs binnen vijf minuten variëren.
- De meeste hits van zoekmachines blijken irrelevant waardoor het totaal aantal hits een betekenisloos getal is. Ook is onzeker of de relevante hits wel op de eerste resultaatpagina's terechtkomen.
- Daardoor is het in de praktijk onmogelijk de resultaten werkelijk te screenen.

- Het gebruik van zoekmachines voor het aanleggen van een dataverzameling is bijzonder arbeidsintensief, tenzij de onderzoeker specifieke software benut voor het geautomatiseerd downloaden en opslaan van de relevante hits.

Onderzoek dat zoekmachines gebruikt zonder zich expliciet met bovenstaande problemen en beperkingen te verstaan is methodologisch dan ook kwetsbaar. Wie een trend in de tijd ontdekt in een met Altavista verzamelde dataset, kan op basis daarvan nog geen uitspraken doen over een trend 'op het web' (Leydesdorff & Curran 2000). We willen echter niet stellen dat het gebruik van zoekmachines taboe is. Zorgvuldig gebruik in samenhang met andere methoden van dataverzameling kan bijzonder goed uitpakken. De beste resultaten worden doorgaans verkregen door de data te vergelijken en aan te vullen met zoekmachine-onafhankelijke data. Toch ziet het er naar uit dat de toekomst van het verzamelen van data voor sociaal-wetenschappelijk interessant internetonderzoek niet ligt in het gebruik van commerciële zoekmachines. Een aantal onderzoeksgroepen heeft inmiddels besloten zelf software te ontwikkelen voor een meer transparante vorm van dataverzameling op het web en op het internet (Thelwall 2001f; Neustadt, Robinson & Kestnbaum 2002; Foot & Schneider 2003; White, Carey & Dailey 2003). Zoals bekend beperken zoekmachines zich tot het publiek toegankelijke web. Daarbinnen is hun reikwijdte bovendien beperkt door de robotethiek.¹⁸ Meer gespecialiseerde software kan het mogelijk maken ook de andere internetprotocollen te benutten en databases (half)automatisch te doorzoeken. Deze ontwikkeling staat anno 2003 echter nog in de kinderschoenen. Indien deze trend zich doorzet, zullen hogere eisen worden gesteld aan de kwaliteit van de software die de onderzoeker gebruikt. Het stadium van dataverzameling in sociaal-wetenschappelijk onderzoek zal dan in toenemende mate afhankelijk worden van de beschikbaarheid van adequate technische en personele infrastructuur. Dan kan immers niet meer worden verwacht dat de individuele onderzoeker zelf de softwareontwikkeling aandrijft. Internetonderzoek draagt dan met andere woorden bij aan het kapitaalintensiever worden van de sociale wetenschappen.¹⁹

18 <http://www.robotstxt.org/wc/norobots.html>.

19 Zie voor een verhandeling over de relatie tussen ICT en de onderzoeksagenda van de geesteswetenschappen Bijker & Peperkamp 2002.

Het benutten van internetcommunicatie voor het aanleggen van een dataverzameling

We hebben hierboven al aangegeven dat de toenemende beschikbaarheid van websites van onderzoekers en onderzoeksgroepen het soms gemakkelijker maakt relevante, reeds geschapen dataverzamelingen op te sporen. Hier willen we ingaan op het on line-interview en de on line-enquête (ofwel het gestandaardiseerde interview). Met beide vormen is inmiddels ervaring opgedaan. De onderzoeker zal zich uiteraard in de eerste plaats baseren op de handboeken over het face to face-interview. Bij het houden van een enquête blijven de grondregels van enquêtes in principe eveneens geldig. Het gaat hier immers niet om een radicaal nieuwe vorm maar om een nieuw type mediatie, dat sommige aspecten van het interview en de enquête betreft, maar niet alle (Mann & Stewart 2000).

De belangrijkste nieuwe aspecten van on line interviews en enquêteren zijn:

- Het afwezig zijn van fysieke aanwijzingen betreffende identiteit en gedrag van de respondent (gender, maatschappelijke status, gemoedsgesteldheid, aandacht, begrip van de vraag). Dit kan overigens bevorderend zijn voor het stem geven aan minderheden en onderdrukte groepen. De afwezigheid van fysieke aanwijzingen heeft consequenties voor de wijze waarop de onderzoeker de medewerking van de respondenten kan bevorderen en handhaven. Bovendien kan de authenticiteit van de data in gevaar komen.
- De kostenstructuur van het onderzoek (minder tijd en geld nodig voor reizen, meer tijd en geld nodig voor ICT, verwerkingkosten). Het scheppen van een geschikte fysieke ruimte waarin de respondent deelneemt aan het onderzoek is overigens ook in het on line-onderzoek een belangrijk aandachtspunt. Wellicht nog meer dan in het face to face-werk, doordat de onderzoeker doorgaans niet ter plekke aanwezig is. Ook is zorg voor het handhaven van internettoegang voor de respondenten noodzakelijk, hetgeen vooral bij langlopend onderzoek problematisch kan blijken.
- Selectie van de respondenten en toegangscriteria tot het onderzoek (de respondent moet internettoegang hebben; internetpubliek is doorgaans niet representatief voor het totale publiek). Bij enquêtes die op een publiek toegankelijke website zijn geplaatst heeft de onderzoeker geen controle over het zelfselectieproces van respondenten. Bovendien dient gewaakt te worden voor meervoudige indiening door respondenten. Indien de respondent een andere computer gebruikt is meervoudige indiening niet te detecteren. Een oplossing voor dit probleem is het beperken van de toegang met be-

hulp van unieke gebruikersnamen en wachtwoorden en het benutten van scripts die de toegang tot de enquête na de respons afsluiten.

- Respons rate en het patroon van de respons. De resultaten tot nog toe zijn niet geheel eensluidend. Sommige onderzoekers zijn enthousiast over de mogelijkheid in enkele dagen duizenden antwoordformulieren te verkrijgen. Anderen hebben juist een verlaagde respons rate ervaren. Veel hangt af (net als in face to face-enquêtes overigens) van de wijze van presenteren, de mate waarin respondenten feedback en herinneringen krijgen, en het respecteren van de netiquette. Het is bijvoorbeeld *not done* om ongevraagd enquêteformulieren te e-mailen (O'Connor & Madge (2001) werden onmiddellijk geflamed toen zij dit toch deden).
- Verwerking van de data. On line-interviews zijn onmiddellijk in zijn geheel beschikbaar waardoor het tijdrovende proces van transcriptie overgeslagen kan worden. Hoe belangrijk dit voordeel is hangt overigens af van de verbreiding van spraakherkenningssoftware voor het automatisch inlezen van interviews van ingesproken tapes of discs. Voor on line-surveys is inmiddels een aantal pakketten beschikbaar waardoor de eerste fase van de verwerking automatisch kan geschieden.

Het on line-interview kan synchroon (kletskanaal) en asynchroon (e-mail) plaatsvinden. Voor on line-groepsinterviews is de methodologie van de focusgroep het beste uitgangspunt. Enquêtes kunnen op een website worden geplaatst, per e-mail worden verspreid en worden gecombineerd. We kunnen in dit bestek niet in detail ingaan op de wijze waarop de internetonderzoeker het on line-interview kan organiseren.²⁰ Behalve de al genoemde problemen, zijn de belangrijkste aandachtspunten:

- Het opbouwen en onderhouden van de vertrouwensrelatie. Dit vereist een grote mate van discursiviteit van de kant van de onderzoeker. De respondent zal bijvoorbeeld de wetenschappelijke reputatie van de onderzoeker on line willen controleren en inzage willen in de doelstelling en opzet van het onderzoek. Ook is het raadzaam elke stap in het interview uitgebreid toe te lichten. In synchrone groepsinterviews is veel aandacht vereist voor de groepsdynamiek. Gezien de arbeidsintensiviteit van het managen van een on line-focusgroep is het inzetten van minimaal twee onderzoekers (met ieder een eigen rol) geen overbodige luxe.

20 Zie hiervoor Mann & Stewart 2000.

- Bescherming van de privacy. Internetcommunicatie laat meer dan analoge communicatie digitale sporen achter. Gebruik hiervan dient in overeenstemming te zijn met de nationale wetgeving en de beroepscodes van de betreffende discipline. Voor internetonderzoek zijn de ethische richtlijnen van de AoIR <http://www.AoIR.org/reports/ethics.pdf> bovendien relevant.
- Gevoeligheid van de gevraagde data. Het relatief anonieme karakter van internetcommunicatie kan het zowel eenvoudiger als ingewikkelder maken om data over gedragsaspecten die als gevoelig worden ervaren te verkrijgen.
- Respect voor de lokale netiquette. Er bestaan uiteraard geen uniforme gedragsregels op het internet. E-mail, kletskanalen, nieuwsgroepen, en on lineconferenties hebben hun specifieke regels ontwikkeld. Bovendien kent iedere e-mailgroep een lokale cultuur waarop de onderzoeker niet ongestraft inbreuk kan plegen. Hieronder valt ook taalgebruik en het hanteren van emoticons en andere paralinguïstische technieken.
- Sampling technieken. Zoals gezegd is de on linepopulatie als zodanig onbekend. Data-archieven kunnen wel nuttig zijn bij het opsporen van relevante on linedoelgroepen. Het voordeel van on linesampling is dat de onderzoeker de hele on linewereld tot haar beschikking heeft. Dit maakt het mogelijk verspreide subculturen te onderzoeken die zonder het internet niet of nauwelijks in kaart zijn te brengen of vergelijkend onderzoek te doen naar groepen in verschillende delen van de wereld.
- Een zorgvuldige keuze van de software. Elk pakket vergemakkelijkt sommige vormen van respons op het interview en de enquête, maar maakt andere gedragspatronen lastiger. Dit kan een sterk sturende invloed hebben op de doelgroep. Ook het ontwerp van de interface met de gebruiker is cruciaal voor de wijze waarop de respondent reageert.

Internet en indicatorconstructie

Het bouwen van internetindicatoren heeft zich tot nog toe voornamelijk ontwikkeld in het commerciële domein rondom websites en on line adverteren. De constructie van indicatoren voor sociaal-wetenschappelijk gebied bevindt zich nog in een beginfase. De aard van de internetdata en de structuur van de software gebruikt voor het aanleggen van de dataverzamelingen hebben een sterke invloed op de constructie van indicatoren voor kwantitatief sociometrisch, bibliometrisch, infometrisch, econometrisch en webometrisch onderzoek. In het bouwen van indicatoren zijn twee eigenschappen cruciaal, die beide door het gebruik van internet data worden beïnvloed:

- Validiteit: de mate waarin de indicator het te meten fenomeen of proces representeert;

- **Betrouwbaarheid:** de mate waarin de indicator het proces op dezelfde wijze meet.

Het zal duidelijk zijn dat de hierboven genoemde problemen met internetdata onmiddellijk gevolgen hebben voor zowel de validiteit als de betrouwbaarheid van de data. Gezien de stand van zaken in kwantitatief internetonderzoek zijn er nog niet veel routines waarop de onderzoeker kan vertrouwen. Het vaststellen van de validiteit van indicatoren is afhankelijk van het kwalitatieve onderzoek naar de wijze waarop internetgebruik maatschappelijk functioneert. Pas op basis daarvan kan immers de relatie worden vastgesteld tussen bepaalde internetindicatoren en de processen waarin de onderzoeker is geïnteresseerd. Het is dan ook aan te bevelen te onderzoeken of dergelijk onderzoek al is verricht. Indien dat niet het geval is zal een fase van exploratief kwalitatief onderzoek ingelast kunnen worden.²¹ De betrouwbaarheid van internetindicatoren wordt direct bedreigd door de instabiliteit van het internet. Tenslotte kan deze instabiliteit ertoe leiden dat sommige indicatoren gedurende een bepaalde, onvoorspelbare, tijd simpelweg onmeetbaar blijken. Deze problemen zijn overigens niet voor alle internetdata even belangrijk. Het is wel degelijk mogelijk robuuste dataverzamelingen aan te leggen.

Zijn deze formidabele obstakels genomen, dan ziet de kwantitatief onderzoeker uit op een luilekkerland van data. Karakteristiek voor het internet is immers dat vrijwel alle activiteit digitale sporen nalaat die stuk voor stuk zijn te vangen in datasets. Belemmeringen liggen dan eerder op het vlak van de ethiek en de werkbaarheid dan op het technologische vlak. Uiteraard bestaat het gevaar dat de onderzoeker te veel wordt gedreven door de data, in plaats van door interessante vragen, maar dat is als zodanig niet specifiek voor het internet.

Data-analyse en -interpretatie

Op het gebied van de interpretatie van de onderzoeksgegevens bestaan voor zover bekend geen specifieke uitdagingen of problemen vanwege het gebruik van het internet in het onderzoek. Op het gebied van de analyse ligt de innovatie in sociaal-wetenschappelijke methodologie eerder in het gebruik van geavanceerde softwarepakketten (SPSS, UCINET, Simulatiesoftware, Atlas,

²¹ Dat hebben we bijvoorbeeld gedaan in ons Wiser-project (<http://www.webindicators.org/>).

Nudist/NVivo, Netwerkanalyse, Discoursanalyse, computerondersteund interviewen, Survey analysesoftware) dan dat het internetgebruik zelf zoveel verschil zou uitmaken. Wel vinden er interessante ontwikkelingen plaats op het kruisvlak van dataverzameling en computerondersteunde analyse. Het gebruik van het internet als communicatiemiddel kan het tenslotte wel een stuk eenvoudiger maken met collega's elders een gezamenlijke analyse uit te voeren (al of niet met behulp van cscw of online conferencing software). De al eerder genoemde portal webuse²² presenteert analytische software die via de portal kan worden benut (Neustadt et al. 2002). De onderzoeker hoeft deze software dus niet meer zelf op de lokale computer te installeren, maar kan de analyse van data direct via het web uitvoeren.²³ De mogelijkheden van dit type analytische software is tot nog toe wel beperkter dan de traditionele softwarepakketten. In welke mate het over het web uitvoeren van grootscheepse data-analyse in de sociale wetenschappen voordelen biedt, is op dit moment moeilijk te overzien.

In het bestek van dit kookboekje voor de internet onderzoeker willen we kort drie exemplarische ontwikkelingen in de analyse aanstippen:

- De opkomst van virtuele etnografie;
- Webometrie en hyperlink netwerk analyse;
- Grootschalige website analyse.

Virtuele etnografie

De kern van etnografie is het recreëren van de leefwereld zoals die eruitziet vanuit het standpunt van de actor (Hammersley & Atkinson 1983). Doorgaans worden data voor dit onderzoek in veldwerk ter plekke verzameld door middel van langdurig volgehouden observaties of participerende observaties. Dit wordt veelal gecombineerd met data uit interviews. De selectie van de onderzoekssite neemt in etnografie dan ook een centrale plaats in. In virtuele etnografie krijgt het begrip veldwerk een andere betekenis (Hine 2000). Niet een lokale onderzoekssite maar een netwerk van websites staat centraal. Met een of meerdere websites als beginpunt, traceert de etnografie het netwerk zoals de actoren dat ook doen. In het geval van participerende observatie onderneemt de onderzoeker ook virtuele acties. De methodologische grondslagen verschillen overigens niet van face to face-etnografie. Dit geldt ook voor de ethische kanten. Het verschil tussen off line en on line etnografisch werk is in grote lijnen identiek

22 <http://www.webuse.org>.

23 Een vergelijkbaar instrument is de General Inquirer van Harvard University <http://www.wjh.harvard.edu/~inquirer>.

aan het hierboven besproken verschil tussen off line en on line interviewen. Virtuele etnografie vertoont meer overlap met discoursanalyse dan zijn reële tegenhanger doordat websiteanalyse een belangrijke plaats inneemt, althans zoals deze zich tot nog toe heeft ontwikkeld. De combinatie van virtuele en face to face-etnografie lijkt bijzonder veelbelovend (Beaulieu 2001). In de analyse van de verkregen data (die nogal omvangrijk kunnen uitpakken) zijn softwarepakketten als Atlas en Nudist bijzonder bruikbaar. Evenals off line-etnografie is virtuele etnografie bijzonder tijdrovend.

Webometrie en hyperlink netwerkanalyse

De grote hoeveelheid kwantitatieve data op het internet hebben onderzoekers uit verschillende disciplines (fysica, economie, informatiewetenschap, bibliometrie, modelleren) ertoe gebracht onderzoek te doen naar regelmatigheden. De samenvloeiing van deze verschillende benaderingen en perspectieven wordt webometrie genoemd (Almind & Ingwersen 1997; Björneborn & Ingwersen 2001; Bar-Ilan & Peritz 2002).

Onder deze noemer worden verschillende typen onderzoeksvragen gesteld. Zo wordt het internet gezien als een model van een zelforganiserend systeem (Barabási 2001, 2002). Andere onderzoekers zijn vooral geïnteresseerd in het herkennen van off line processen in on linedata. Methodologisch zit het specifieke vooral in de hierboven beschreven eigenschappen van internetdata. Een mogelijke analytische innovatie kan de ontwikkeling zijn van genetwerkte simulaties en modelleringen, waardoor simulaties mogelijk een complexer geheel kunnen representeren (Ahrweiler & Gilbert 1998; Gilbert & Troitzsch 1999; Gilbert 1997; Leydesdorff 1995; Leydesdorff 2002). In de tweede plaats is de combinatie van webometrie met hyperlink netwerkanalyse (Park, Barnett & Nam 2002) veelbelovend. Hyperlink netwerkanalyse ontleent zijn methodologische en theoretische grondslag aan sociale netwerkanalyse, maar gebruikt de hyperlinks tussen websites als representant van sociale verbindingen. Op deze wijze kunnen on line- en off linenetwerken met elkaar worden vergeleken.

Grootschalige website-analyse

Ten slotte levert de combinatie van computerondersteunde inhoudsanalyse met geautomatiseerd downloaden van websites en webpagina's een nieuwe vorm op van grootschalige website-analyse (Foot & Schneider 2003). Door websites in zijn geheel te downloaden en met gebruik van dezelfde software te annoteren komt de mogelijkheid in zicht de kracht van een gedetailleerde inhoudsanalyse van een klein aantal cases te combineren met de sterke kant van een vergelijkende analyse van een groot aantal documenten. Het ligt in de ver-

wachting dat meer onderzoeksgroepen dergelijke onderzoeksinstrumenten zullen ontwikkelen. De perspectieven van deze aanpak zijn nog niet helemaal helder.

Enkele nuttige meta-methodologische principes

In het bestek van dit beknopte kookboekje zijn veel details onbesproken gebleven. De onderzoeker die verder wil gaan, kan deze vinden in de geciteerde literatuur en de genoemde websites. Om deze literatuur goed te kunnen benutten kunnen de volgende theoretische, meta-methodologische, uitgangspunten wellicht behulpzaam zijn:

- Symmetrisch analytische scepsis (Bloor 1991; Woolgar 1988). De onderzoeker doet er wijs aan scepsis met betrekking tot de claims rondom ICT en het internet in te bouwen in de onderzoeksopzet. Dit geldt zowel opgeklopte hoop als doemdenken. Het debat tussen de 'revolutionairen' (Castells 1996) en de 'evolutionairen' (Webster 1995) betreft de vraag of het internet een breuk in de maatschappelijke ontwikkeling representeert. Beide lijken er echter van uit te gaan dat dit een eigenschap is van het object van onderzoek (internet en de samenleving). Het principe van analytische scepsis plaatst dit probleem meer aan de kant van het perspectief van de onderzoeker. Afhankelijk van de specifieke onderzoeksvraag zal het internet een geleidelijke ontwikkeling dan wel een radicale breuk vertegenwoordigen. Doel van het onderzoek is dan niet zozeer het ontmaskeren van de hype, alswel het zien van de hype, en de kritiek daarop, als deel van het object van onderzoek.
- Zie kennis (en maatschappelijke praktijken) altijd als gesitueerde kennis (en praktijken) (Haraway 1991). Dit principe is iets bruikbaar dan het begrip 'lokale kennis' (Geertz 1983) omdat het de mogelijkheid erkent dat kennis niet zozeer op een plaats is verankerd alswel zijn basis heeft in een netwerk. De neiging om het internet als vrije virtuele ruimte te zien (op zichzelf overigens een interessante eigenschap van de communicatie over het internet) kan met dit uitgangspunt systematisch worden weerstaan. Ook nodigt het de onderzoeker uit altijd op zoek te gaan naar de context van productie.
- Benut de mogelijkheden technologie als tekst (Woolgar & Ashmore 1988) en tekst als technologie (Braverman 1974) te analyseren. Vooral in het geval van tekstuele technologie zoals software, hyperlinks en websites kan dit bijzonder vruchtbaar zijn.

- Betrek de constructie van ogenschijnlijk afgeronde maatschappelijke producten en processen in de analyse. Dit heeft in het wetenschaps- en techniekonderzoek bijzonder vruchtbare resultaten afgeworpen. In de analyse van nieuwe verschijnselen zoals het internet ligt het overigens nogal voor de hand.
- Besteed extra aandacht aan processen van mediatie en representatie. Mediatie en representatie zijn niet 'onschuldig' (Thompson 1995). Dit is zeker niet specifiek voor de bestudering van het internet. In het algemeen is het duidelijk dat de massamedia grote invloed hebben op maatschappelijke processen op alle niveaus. Niet in alle sociaal-wetenschappelijke disciplines is deze aandacht echter al sterk ontwikkeld. Wellicht kan de bestudering van het internet hiertoe verder stimuleren. Aan te bevelen naslagwerken, websites en referenties.
- Houd de 'ouderwetse' methodologische uitgangspunten van het eigen vakgebied in ere. Het internet is althans geen reden ze overboord te zetten. Methodologische innovatie is doorgaans meer gebaat bij het benutten van in jarenlange onderzoekservaring ontwikkelde principes dan in het negeren of afleren ervan. In het laatste geval zou multi- en interdisciplinair werk overigens zijn betekenis verliezen.

Ter afsluiting

Dit laatste punt, houd de discipline-gebonden methodologische ervaringen in ere, is wellicht de centrale boodschap van dit kookboekje. Het is een wat paradoxale. Het gebruik van het internet in sociaal-wetenschappelijk onderzoek kan inderdaad tot belangrijke innovatie in methoden en stijl van het onderzoek leiden. Dat wordt echter niet bevorderd door al te veel enthousiasme over het web en het internet. Een sceptische houding, waarin de inbedding van de on line- in de off linewereld een van de kernpunten is, levert eerder een vruchtbaar perspectief op. We staan nog maar aan het begin van een lange weg, waarin in rol en aard van sociaal-wetenschappelijke kennis diepgaande ICT-gerelateerde veranderingen kunnen optreden. In het begaan van die weg hebben we nuchterheid en scepsis hard nodig.

- Adamic, L.A. & B.A. Huberman (2000) 'Scaling behaviour of the world wide web', *Science*, 287: 2115.
- Ahrweiler, P. & N. Gilbert (1998) *Computer Simulations in Science and Technology Studies*, Berlin, Heidelberg, New York: Springer Verlag.
- Almind, T. & P. Ingwersen (1997) Informetric analyses on the world wide web: methodological approaches to 'webometrics', *Journal of Documentation*, 53: 404-426.
- Bar-Ilan, J. & B.C. Peritz (2002) Informetric theories and methods for exploring the Internet: an analytical survey of recent research literature, *Library Trends*, 50 (3): 371-392.
- Bar-Ilan, J. (2001) 'Data collection methods on the web for informetric purposes: a review and analysis', *Scientometrics*, 50 (1): 7-32.
- Barabási, A.-L. (2001) 'The physics of the web', *Physics World*, July, www.physicsweb.org/article/world/14-7/09.
- Barabási, A.-L. (2002) *Linked. The New Science of Networks*, Cambridge, Mass.: Perseus Publishing.
- Beaulieu, A. (2001) 'Combining users, logs and texts: tracing networks of trust in data-sharing', *Virtual Methods Seminar*.
- Berners-Lee, T. (1999) *Weaving the Web. The Original Design and Ultimate Destiny of the world wide web by its Inventor*, New York: Harper-Collins Publishers.
- Bijker, W. & B. Peperkamp (2002) (Report No. Achtergrondstudie nr. 27) Den Haag: Adviesraad voor het Wetenschaps- en Technologiebeleid.
- Björneborn, L. & P. Ingwersen (2001) 'Perspectives of webometrics', *Scientometrics*, 50 (1): 65-82.
- Bloor, D. (1991) *Knowledge and Social Imagery* (2nd ed.), Chicago: University of Chicago Press.
- Braverman, H. (1974) *Labor and Monopoly Capital*, New York, London: Monthly Review Press.
- Carpenter, L., S. Shaw & A. Prescott (1998) *Towards the Digital Library. The British Library's Initiatives for Access Programme*, London: The British Library.
- Castells, M. (1996) *The Rise of the Network Society*, Cambridge, Mass.: Blackwell Publishers.
- Davenport, E. & H. Hall, (2002) 'Organizational knowledge and communities of practice', in B. Cronin & D. Shaw (eds.), *Annual Review of Information Science and Technology*, pp. 171-227, Medford, New Jersey: Information Today, Inc.

- Davies, P.H.J. (2002) 'Intelligence, information technology, and information warfare', in B. Cronin (ed.), *Annual Review of Information Science and Technology*, 36: 313-352, Information Today Inc.
- Dutton, W.H. (1996) *Information and Communication Technologies. Visions and Realities*, Oxford: Oxford University Press.
- Foot, K. & S. Schneider (2003) 'Alliances or antagonies? Hyperlinks & associative relations in web sphere analysis', *Journal of Computer-Mediated Communication*.
- Fox, E.A. & S.R. Urs (2002) 'Digital libraries', in B. Cronin (ed.), *Annual Review of Information Science and Technology*, Medford, NJ: Information Today Inc.
- Garton, L., C. Haythornthwaite & B. Wellman (1997) 'Studying online social networks', *Journal of Computer-Mediated Communication*, 3 (1).
- Geertz, C. (1983) *Local Knowledge*, New York: Basic Books.
- Gilbert, G.N. (1997) 'A simulation of the structure of academic science', *Sociological Research Online*, 2 (2): ppl-2.
- Gilbert, N. & K.G. Troitzsch (1999) *Simulation for the Social Scientist*, Buckingham, Philadelphia: Open University Press.
- Ginsparg, Paul (1996) 'Winners and Losers in the Global Research Village', *The Serials Librarian*, 30 (3/4): 83-90.
- Hammersley, M. & P. Atkinson (1983) *Ethnography. Principles in Practice*, London: Routledge.
- Haraway, D. (1991) *Simians, Cyborgs and Women. The Reinvention of Nature*, New York: Routledge.
- Harter, S.P. (1997) 'Scholarly communication and the digital library: problems and issues', *Journal of Digital Information*, 1 (1), (<http://jodi.ecs.soton.ac.uk/Articles/v01/i01/Harter/>).
- Hine, C. (2000) *Virtual Ethnography*, London: Sage.
- How Much Information?* (2000) retrieved from <http://www.sims.berkeley.edu/how-much-info> on 20 February 2003.
- Huberman, B.A. & L.A. Adamic (1999) 'Growth dynamics of the World-Wide web', *Nature*, 401: 131, cond-mat/9901071.
- Huizer, E. (2002) *Het internet, een studie waard*, Enschede: Universiteit Twente.
- Introna, L.D. & H. Nissenbaum (2000) 'Shaping the web: why the politics of search engines matters', *The Information Society*, 16: 169-185.
- Jankowski, N.W. (2001) 'Creating community with media: history, theories and scientific investigations', in N.W. Jankowski & O. Prehn (eds.), *Community Media in the Information Age. Perspectives and Prospects*, Cresskill, NJ: Hampton.
- Katz, J.S. & B.R. Martin (1997) 'What is research collaboration?', *Research Policy*, 26: 1-18.

- Kling, R. & G. McKim (2000) 'Not just a matter of time: field differences and the shaping of electronic media in supporting scientific communication', *Journal of the American Society for Information Science*, 51 (14): 1306-1320.
- Koehler, W. (1999) 'Digital libraries and world wide web sites and page persistence', *Information Research*, 4 (4), <http://InformationR.net/ir/4-4/-paper60.html>.
- Koslow, S.H. (2002) 'Sharing primary data: a threat or asset to discovery?', *Nature Reviews Neuroscience*, 3 (4): 311-313.
- Landow, G.P. (1992) *Hypertext 2.0. The Convergence of Contemporary Critical Theory and Technology*, Baltimore: Johns Hopkins University Press.
- Lawrence, S. & C.L. Giles (1999a) 'Accessibility of information on the web', *Nature*, 400: 107-109.
- Lawrence, S. & C.L. Giles (1999b) 'Searching the world wide web: general and scientific information access', *IEEE Communications*, 37 (1): 116-122.
- Lawrence, S., D.M. Pennock, G.W. Flake, R. Krovetz, F.M. Coetzee, E. Glover, F.Å. Nielsen, A. Kruger & C.L. Giles (2001) 'Persistence of web references in scientific research', *Computer*, 34 (2): 26-31.
- Leydesdorff, L. (1995) *The Challenge of Scientometrics. The Development, Measurement, and Self-Organization of Scientific Communications*, Leiden: dsw Press, Leiden University.
- Leydesdorff, L. (2002) 'The communication turn in the theory of social systems', *Systems Research and Behavioral Science*, 19 (2): 129-136.
- Leydesdorff, L. & M. Curran (2000) Mapping university-industry-government relations on the Internet: an exploration of indicators for a knowledge-based economy, *Cybermetrics*, 4 (1).
- Lunn, R.J. & M.W. Suman (2002) 'Experience and trust in online shopping', in B. Wellman & C. Haythorntwaite (eds.), *The Internet in Everyday Life*, pp. 549-577, Malden, Mass.: Blackwell Publishers Ltd.
- Mann, C. & F. Stewart (2000) *Internet Communication and Qualitative Research. A Handbook for Researching Online*, Thousand Oaks, Cal.: Sage.
- Marres, N. & R. Rogers (2000) 'Depluralising the web, repluralising public debate: the case of the GM food debate on the web', in R. Rogers, *Preferred Placement. Knowledge Politics on the Web*, pp. 113-136, Maastricht: Jan van Eyck Editions.
- Miller, D. & D. Slater (2000) *The Internet. An Ethnographic Approach*, Oxford, New York: Berg.
- Molyneux, R.E. & R.V. Williams (1999) 'Measuring the Internet', in M.E. Williams (ed.), *Annual Review of Information Science and Technology*, 34: 287-339, Medford, New Jersey: Information Today Inc.
- Mortensen, T. & J. Walker (2002) 'Blogging thoughts: personal publication as an online research tool', *Intermedia Researching ICTs in Context* (Vol. 3) Oslo, Norway: University of Oslo.

- Nelson, T.H. (1981) *Literary Machines*, Swarthmore, PA.
- Neustadt, A., J. P. Robinson & M. Kestnbaum (2002) 'Doing social science research online', in B. Wellman & C. Haythornthwaite (eds.), *The Internet in Everyday Life*, pp. 186-211, Malden, USA: Blackwell Publishers Ltd.
- Newman, M.E.J. (2000) 'The structure of scientific collaboration networks', *Proceedings of the National Academy of Sciences*, 98 (2): 404-409.
- O'Connor, H. & C. Madge (2001) 'Cybermothers: online synchronous interviewing using conference software', *Sociological Research Online*, 5 (4).
- OECD (1998) *The Global Research Village*, Paris: OECD.
- Olson, G., T. Finholt & S. Teasley (2000) 'Behavioral aspects of collaboratories', in S. Koslow & M. Huerta, pp. 1-14, Mahwah, NJ: Lawrence Erlbaum Associates.
- Park, H.W., G.A. Barnett & I.Y. Nam (2002) 'Hyperlink-affiliation network structure of top websites: examining affiliates with hyperlink in Korea', *Journal of the American Society for Information Science and Technology*, 53 (7): 592-601.
- Rheingold, H. (1994) *The Virtual Community. Homesteading on the Electronic Frontier*, New York: Harper Perennial.
- Schroeder, P. (2001) *The Global Research Village. Conference Report, Access to Publicly Financed Research*, Amsterdam.
- Sherman, C. & G. Price (2001) *The Invisible Web. Uncovering Information Sources Search Engines can't See*, Medford, New Jersey: Information Today Inc.
- Slevin, J. (2000) *The Internet and Society*, Cambridge: Polity Press.
- Steinmueller, W.E. (2002) 'Virtual communities and the new economy', in R. Mansell (ed.), *Inside The Communications Revolution. Evolving Patterns of Social and Technical Interaction*, pp. 21-54, Oxford: Oxford University Press.
- Sudweeks, F. & M. Allbritton (1996) 'Working together apart: communication and collaboration in a networked group', in C.D. Keen, C. Urquhart & J. Lamp, *Proceedings of the 7th Australian Conference of Information Systems. Working together apart: Communication and Collaboration in a Networked Group*, 2: 701-712.
- Sudweeks, F. & S.J. Simoff (1998) 'Complementary explorative data analysis: the reconciliation of quantitative and qualitative principles', S. Jones (ed.), *Doing Internet Research. Critical Issues and Methods for Examining the Net*, London: Sage.
- Taubes, G. (1996) SPECIAL NEWS REPORT: Electronic preprints point the way to 'author empowerment', *Science*, 271 (5250): 767.
- Thelwall, M. (2001f) 'A web crawler design for data mining', *Journal of Information Science*, 27 (5): 319-325.
- Thompson, J.B. (1995) *The Media and Modernity*, Cambridge: Polity Press.

- Van den Besselaar, P., M. Tanabe & T. Ishida (2002) 'Introduction: digital cities research and open issues', *Digital Cities II. Computational and Sociological Approaches. Lecture Notes in Computer Science*, 2362 Berlin: Springer.
- Voorbij, H. (1999) 'Searching scientific information on the Internet: a Dutch academic user survey', *Journal of the American Society for Information Science*, 50 (7): 598-615.
- Wakeford, N. (2000) 'New media, new methodologies: studying the web', in D. Gauntlett (ed.), *Web.Studies. Rewiring Media Studies for the Digital Age*, pp. 31-43, London: Arnold Publishers.
- Walsh, J.P. & N.G. Maloney, 'Computer network use, collaboration structures and productivity', in P. Hinds & S. Kiesler, *Distributed work*, Cambridge, MA: MIT Press.
- Watts, D.J. (1999) *Small Worlds. The Dynamics of Networks between Order and Randomness*, Princeton, New Jersey: Princeton University Press.
- Webster, F. (1995) *Theories of the Information Society*, New York: Routledge.
- Wellman, B. & C. Haythorntwaite (2002) *The Internet in Everyday Life* (The Information Age Series) Malden, USA: Blackwell Publishers Ltd.
- White, J.A., L.M. Carey & K.A. Dailey (2003) 'Web-Based Instrumentation in educational survey research', *Webnet Journal: Internet Technologies, Applications and Issues* (in press).
- Woolgar, S. & M. Ashmore (1988) *Knowledge and Reflexivity. New Frontiers in the Sociology of Knowledge*, London: Sage.
- Woolgar, S. (2002) *Virtual Society? Technology, Cyberbole, Reality*, Oxford: Oxford University Press.
- Wouters, P. & P. Schröder (2000) *Access to Publicly-Financed Research. The Global Research Village III*, Amsterdam: Organisation for Economic Co-operation and Development, Netherlands Ministry of Education, Culture and Science (Min. OCW) NIWI-KNAW.
- Wouters, P. (2002) *Policies on Digital Research Data. An International Survey* (The Public Domain of Digital Research Data), Amsterdam: NIWI-KNAW.
- Wouters, P. & A. Beaulieu (2002) 'Quality control of data in data-sharing practices and regulations', *18th International Conference, CODATA 2002 Frontiers of Scientific and Technical Data*.
- Wouters, P. & P. Schröder (2003) *Promise and Practice in Data Sharing* (The Public Domain of Digital Research Data), Amsterdam: NIWI-KNAW.
- Wyatt, S., F. Henwood, N. Miller & P. Senker (2000) *Technology and In/equality. Questioning the Information Society*, London: Routledge.
- Wyatt, S., G. Thomas & T. Terranova (2002) 'They came, they surfed, they went back to the beach: conceptualizing use and non-use of the Internet', in S. Woolgar (ed.), *Virtual Society? Technology, Cyberbole, Reality*, pp. 23-40, Oxford: Oxford University Press.