

Kvalitetsmålinger ved hjelp av paradata.

Øyvind Kleven¹ og Gustav Haraldsen²
Statistisk sentralbyrå, Norge

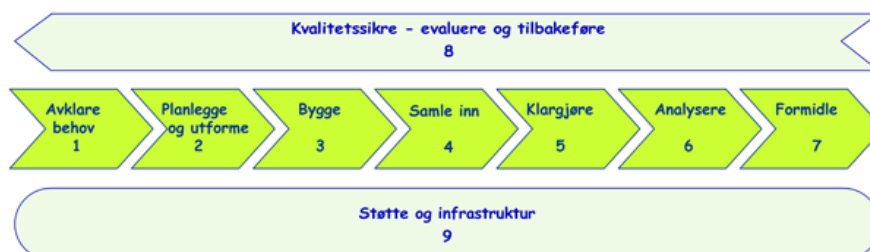
I elektroniske skjemaer er det mulig å legge in applikasjoner som registrerer hvordan respondentene besvarer skjemaet. Antall feilmeldinger som aktiviseres, tiden respondenten bruker, hvilken rute respondenten velger også videre er eksempler på det som i internasjonal metodelitteratur kalles paradata. I Norge har vi gjennom flere år registrert paradata både i intervjuundersøkelser og selvadministrerte skjema og benyttet dem i kvalitetsmålinger. Ved hjelp av vårt Eyetracker utstyr har vi også både kunnet foreta tradisjonelle paradatamålinger og utvidet bruken av paradata i planleggingen av nye datainnsamlinger. Paperet gir en enkel innføring i hva paradata er og hvordan det benyttes i kvalitetsmålinger.

Nøkkelord: paradata, kvalitetsmålinger, brukerkervennlighetstester

Målinger av prosesskvalitet

I statistiske sentralbyråer verden over har man innført en såkalt Generic Statistical Business Process Modell (UNECE 2009). I statistisk sentralbyrå i Norge kaller vi dette vår virksomhetsmodell. Virksomhetsmodellen er innført som et overordnet rammeverk for å beskrive alle prosessene i statistikkproduksjonen (se figur 1). Modellen er et viktig redskap i planleggingen av nye statistikker, og ved arbeid med forbedring og standardisering av eksisterende arbeidsprosesser i statistikkproduksjonen. I kontinuerlig forbedring ved å gjennomføre kvalitetskontroller for hver enkelt prosess er den viktig som et overordnet kart over prosessene (Køber et al. 2008). Ideen bak den er at gode prosesser leder til gode resultater.

Figur 1 Virksomhetsmodell for SSB.



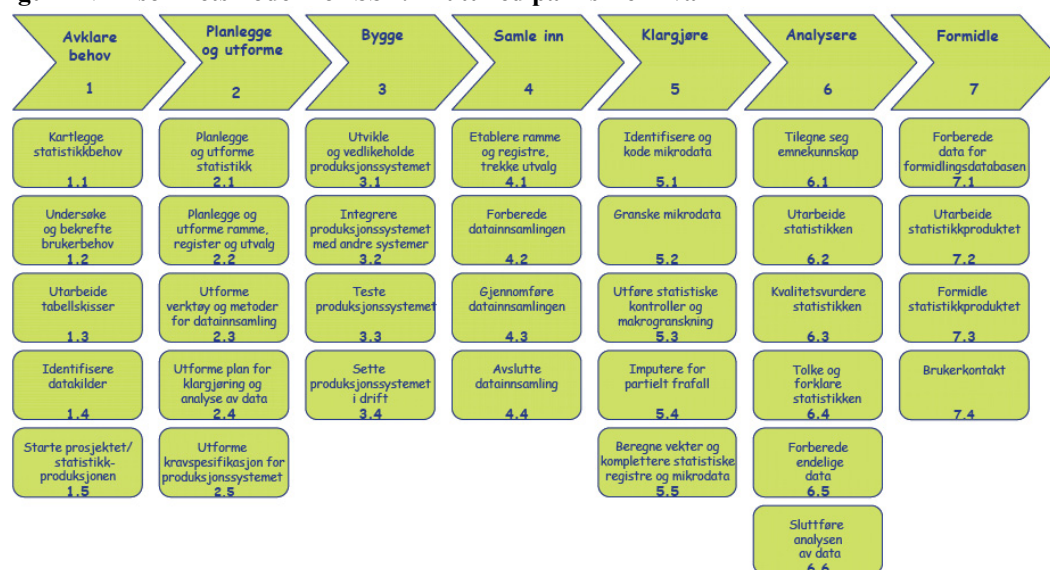
I statistiske sentralbyråer er det viktig å måle kvaliteten i det vi gjør, i virksomhetsmodellen er dette markert ved at det er innført en prosess 8 som heter Kvalitetssikre – evaluere og tilbakeføre som løper overordnet og gjennom alle de andre delprosessene. Hensikten er å forbedre statistikk og analyser i samsvar med brukerbehovene, samt å effektivisere produksjonen. Virksomhetsmodellen beskriver i detalj en rekke prosesser som foregår inne i statistikkkontoret (se figur 2), og tanken er at gjennom å

¹ Øyvind Kleven, Statistisk sentralbyrå, Seksjon for planlegging og brukertesting. Kongens gate 6. E-mail: KLE@SSB.NO

² Gustav Haraldsen, Statistisk sentralbyrå, Avdeling for datafangst. Oterveien 23. E-mail: GHA@SSB.NO

forbedre prosessene vil kvaliteten på sluttproduktet bli bedre. Vi kan derfor skille mellom å forsøke å måle produktkvalitet (relevans, nøyaktighet og pålitelighet, aktualitet og punktelighet, sammenheng og sammenlignbarhet, tilgjengelighet og klarhet) og måling av prosesskvalitet (gode metoder, egnede statistiske prosedyrer, rimelige krav til oppgavegiverne, kostnadseffektivitet). Modellen gir en oversikt over prosessene i en statistikkproduksjon; beskriver hvilke prosesser som inngår men ikke hvordan prosesskvaliteten skal måles. Dette må også bestemmes og det pågår viktige arbeider med dette i flere statistikkbyråer (se for eksempel Haraldsen et al. 2010). Spesielt er det vanskelig å måle kvaliteten i prosess 4.3 fordi den foregår utenfor institusjonen, i møtet mellom intervjuer og intervjuobjekt eller ute hos respondenten som besvarer et selvutfyllingsskjema. Samtidig er de der disse prosessene som i høy grad bestemmer kvaliteten i datainnsamlingen og dermed legger forutsetningene for og eller dårlig statistikk. I tillegg betraktes gode svarprosesser gjerne som et gode i seg selv, fordi et skjema som er lett og greit å besvare har lavere oppgavebyrde enn et skjema som respondenten strever med

Figur 2 Virksomhetsmodell for SSB. Brutt ned på 2-siffernivå



I metodelitteraturen har det ofte blitt uttrykt bekymring for den begrensede kunnskapen som er tilgjengelig om hvordan respondenter svarer på et spørsmål (se for eksempel Sudman, Bradburn og Schwarz 1996:112). Vi har relativt lite kunnskap om det som foregår når en respondent fyller ut et skjema. Det er vanskelig å vite om skjemaet har vært relativt greit å svare på for respondenten uten at det er en person til stede ved utfylling. Det er imidlertid, som tidligere nevnt, vanskelig å måle kvaliteten ”ute hos respondentene”. Siden vi har liten kunnskap og kontroll med datainnsamlingsfasen (4.3) forsøker vi å gjennomføre gode planleggingsprosesser ved å benytte kvalitative testmetoder under prosess 2.3 eller pilottester i 3.3. Siden 1980 tallet har det vært vanlig å bruke kognitive intervjuer som middel for å få innblikk i svarprosessen ute hos respondenten. En vanlig fremgangsmåte er å spørre respondenten om å fylle ut skjemaet med en testleder tilstede som så spør respondenten om hvordan det var å fylle ut skjemaet.

På tross av sin åpenbare nytteverdi har kognitive intervjuer noen utfordringer og mangler, det er for eksempel arbeidsintensivt og kan være sensitivt for uheldige interaksjonseffekter (Haraldsen et al. 2005, Stern 2008). Det er åpenbart at det er behov for å utvikle metoder som kan gi oss informasjon om prosess 4.3 direkte, og det er behov for å utvikle metoder som kan gi mer direkte adgang til hvordan respondenter fyller ut skjemaer. Paradata er blitt tatt i bruk nettopp med dette for øyet. Det er i det internasjonale metodemiljøet en voksende interesse og oppmerksomhet for paradata, og det har vært mange presentasjoner i på metodekonferanser de siste årene (se for eksempel Haraldsen et al. 2005; Haraldsen et al. 2006; Snijkers og Morren 2010; Heervegh 2002).

Ulike typer paradata

Paradata kan defineres som data som beskriver *hvordan* respondenten fylte ut et spørreskjema i motsetning til *hva* de fylte in. Den tradisjonelle definisjonen av paradata begrenset paradata til å være data om selvutfyllingsprosesser, men det overser at prosessdata samlet inn fra intervjuere, for eksempel i form av atferdskoding eller i debriefingssamtaler også er en form for paradata. Når vi i dag is større grad satser på å kombinere intervju med selvutfylling, vil det også bli mer aktuelt å sammenligne disse to typene svarprosesser. I metodelitteraturen i dag brukes paradata nærmest synonymt med prosessdata. Det er viktig å presisere at det er overgangen fra papirskjema til elektroniske skjema som har gjort det mulig å også samle inn prosessdata i selvadministrerte undersøkelser. Det betyr samtidig at denne typen prosessmålinger først og fremst har verdi i undersøkelser med en høy andel elektroniske skjema. Slike data kan i dag samles inn nærmest automatisk ved bruk av elektroniske skjema. Det forutsetter at det er forhåndsprogrammert en kode i datasystemet. En slik kode kan man laste ned gratis fra Dirk Heerweghs hjemmeside på internett (se Heerwegh 2002; Heerwegh 2003). Paradata kan altså samles inn både i CAI skjemaer og web skjemaer. Denne form for paradata kan deles inn i (Heerwegh 2003) Server Side Paradata og Client Side Paradata.

- Server Side Paradata (SSP) er informasjon som genereres i serveren og dreier seg om hvilke websider brukeren klikker seg inn på.
- Client Side Paradata (CSP) er informasjon om hva som foregår innenfor en webside.

I intervjuundersøkelser kan vi i tillegg til å samle data om svarprosessene få intervjuerne til svare på hvorfor respondenten bruker lang tid på å besvare spørsmål eller endrer allerede avgitte svar. Selv om det er mulig å bygge inn oppfølgingsspørsmål knyttet til paradata i selvutfyllingsprosesser, nøyer vi oss vanligvis med kun å kartlegge atferden. Ved å knytte sammen informasjon fra kvalitative planleggingsmetoder sammen med paradatamålinger kan vi imidlertid foreta kvalifiserte gjetninger om dette (se neste side for et eksempel).

Vi skal også gi et eksempel på utvidelser av paradata. Blikksporing er en type utvidelse av prosessdata som kan brukes i planleggingen av undersøkelser. Hensikten er å utforme skjemaene slik at øynene faller på de informasjonselementene som er viktige og at informasjonselementene er plassert slik at respondentene leser dem i riktig rekkefølge og på en effektiv og grei måte. Første forutsetningen for at respondentene skal tenke ut et korrekt svar på spørsmålene vi stiller er jo at de blir sett og lest.

Kvalitetsmålinger og paradata

Paradata kan være nyttige som indirekte eller direkte målinger av prosess 4.3 i virksomhetsmodellen (gjennomføre datainnsamlingen). Vi kan kalle dette en ekstern prosess siden det er en prosess vi har liten kontroll over. Men det er like fullt en svært viktig prosess siden det er i møtet mellom respondent og spørreskjema kvaliteten på dataene som samles inn avgjøres. Et eksempel kan være illustrerende. I en kundeundersøkelse SSB gjennomførte for noen år siden benyttet vi paradata for å evaluere kvaliteten på spørsmålene (for detaljer se Haraldsen et al. 2005). Vi registrerer tiden hver respondent bruker på å fylle ut hvert skjema og hvor mange som forandrer svaret sitt. Teorien bak er at spørsmål hvor respondentene bruker lang tid på å fylle ut eller de forandrer svaret sitt er det grunn til å kikke nærmere på.

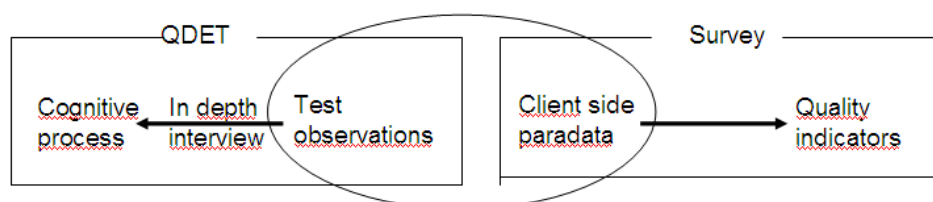
Tabell 1. Thinking about last time you bought something from Statistics Norway, were you satisfied or dissatisfied with the following? Give one answer on each line.

	Very satisfied	Satisfied	Neither satisfied nor dissatisfied	Dissatisfied	Very dissatisfied	Don't know	Not applicable	Average responsetime with and without outliers		Percent who changed their answers
The time it took to get in contact with the right person	42	36	8	4	0	2	8	12	9.1	4.0
The service given	45	41	7	1	0	2	3	2.5	2.1	3.7
Professional competence	42	38	10	1	0	5	4	2.1	1.7	3.3
Statistics Norway's ability to keep deadlines	41	34	12	5	1	1	5	2.6	2.2	5.2
The time it took from first contact to delivery	37	33	14	5	3	1	7	3.4	3.0	3.0
The price you had to pay	12	31	36	10	5	3	2	7,0	4.4	8.1
The product compared with expectations	35	45	13	3	0	2	2	4.3	3.8	4.8
Internal coordination of the service given	23	27	13	5	1	16	16	3.8	3.4	1.8

I tabell 1 se vi fordelingen til 8 spørsmål om hva kundene til SSB synes om ulike ved SSB. Bortest mot høyre ser vi paradata. For det første ser vi at respondentene bruker i gjennomsnitt ganske kort tid på å svare på spørsmålene (mellom 12 – 2 sekunder) og at det er svært få som forandrer svarene sine (mellom 8-2 prosent) Vi ser at respondentene bruker relativt lang tid på å besvare første spørsmål, noe som ikke er merkelig siden de må lese overskriften også. Videre ser vi at når det gjelder spørsmålet om hvor mye en måtte betale så går tiden litt opp og det er relativt flere som forandrer svaret sitt. Forut for selve undersøkelsen gjennomførte vi flere kognitive tester av det potensielle skjemaet, i de testene ble det påvist at spørsmålet om pris var vanskelig å svare på for noen, siden det for eksempel ikke var enkelt å gjøre en sammenlikning siden SSB var den eneste leverandøren. Dette ble dermed bekreftet av paradatene, men på den annen side er det en relativt liten andel som f.eks forandrer spørsmålet sitt. Paradata er kun et monitoringsinstrument det gir ikke nødvendigvis en anbefaling for hvordan vi skal forbedre et skjema.

I SSB har vi tenkt på å benytte paradata som et supplement til tradisjonelle kvalitative tester av skjema, der hvor vi identifiserer problemer i skjemaet i små strategiske tester kan vi etterprøve det ved paradata. Dette er uttrykt i figure 3. Paradata kan benyttes til å kvantifisere kognitive problemer eller for å identifisere spørsmål som bør testes kvalitativt. Paradata kan dermed sies å ”avsløre” dårlige spørsmål og byrdefulle prosesser som ellers er gjemt inne i elektroniske skjemaer.

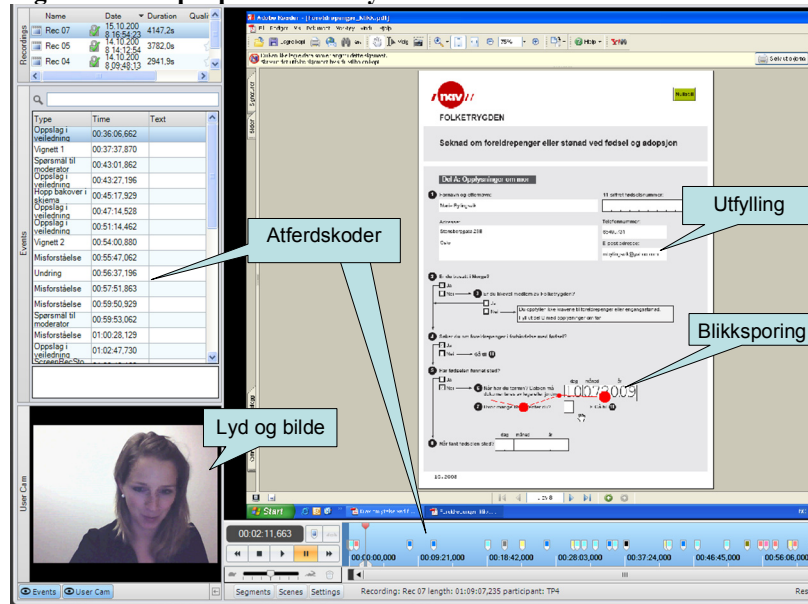
Figure 3: A conceptual model that links together small scale investigations of cognitive processes with big scale evaluations of survey quality



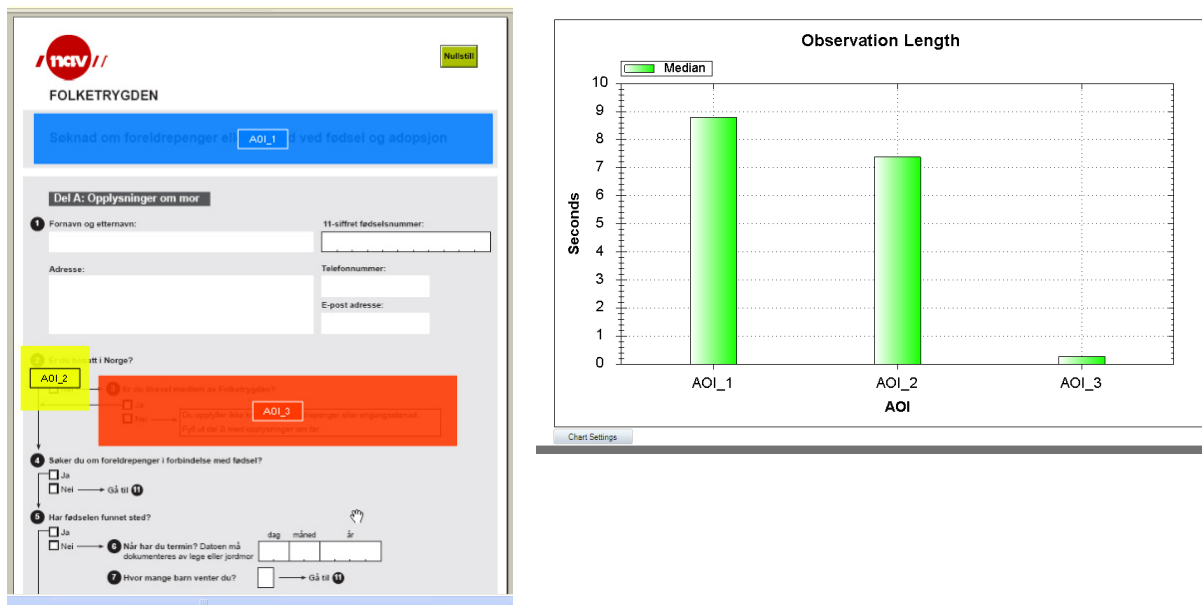
Eyetracker og paradata

Stadig flere statistikkinnsamlere eller offentlige etater som samler inn administrative data benytter nå såkalte eyetrackere for å teste portaler eller elektroniske skjemaer. Eyetrackere har flere bruksområder (se figur 4), det tar opp lyd og bilde av testpersonen mens denne fyller ut skjemaet, den registrerer hvor på skjermen testpersonen ser, hvor lang tid testpersonen bruker på hvert spørsmål registreres automatisk, det er også mulig for testlederen å legge inn adferdskoder direkte inn i systemet.

Figur 4 Eksempel på bruk av Eyetracker



Figur 5 Eksempel på fokusområder



I figur 5 viser vi hvordan bruk av Areas of Interest (AOI) eller fokusområder ser ut. I eyetracker systemet registreres det i sekunder hvor lenge testpersonen ser på ulike spørsmål eller deler av en webside. Det er så mulig å få ut statistikk fra automatisk eller manuelt valgte interesseområder.

I figur 6 ser vi hvordan client side paradata som antall ganger testpersonen har trykket ned en tast eller benyttet musen er registrert, i tillegg har testlederen kodet adferden til testpersonen, for eksempel at det er et vurderingsproblem.

Figur 6 Atferdsregistrering og client side paradata ved Eyetracker

Participant properties	B	C	D	E
17	Atferdsregistrering			
18				
19	Skjema forholds	2 = er av lastebil firma og sjåfør		
20	Alder	3 = > 60 År		
21	Kjøn	1 = mann		
22				
23				
24				
25				
26	Timestamp	Event	Event1	Data1
27	0	InstructionStart	2048	0
28	6797	InstructionEnd	4096	0
29	6909	InstructionStart	2048	0
30	18911	InstructionEnd	4096	0
31	17025	ScreenShotStart	8192	0
32	39392	URL_Start	512	0
33	61728	KeyPress	4	66
34	68223	KeyPress	4	69
35	68845	KeyPress	4	68
36	69091	KeyPress	4	62
37	96038	KeyPress	4	90
38	99208	KeyPress	4	100
39	100481	KeyPress	4	96
40	100203	LeftMouseButton	1	264
41	120828	KeyPress	4	83
42	122965	KeyPress	4	69
43	128412	KeyPress	4	78
44	128893	KeyPress	4	78
45	133822	KeyPress	4	69
46	146538	KeyPress	4	80
47	166427	LeftMouseButton	1	184
48	167738	LogData	0	Vurderingsproblem
49	177760	LeftMouseButton	1	lybe - vet aldri hva et sÅmmelass veier
50	183392	URL_End	1024	0
51	183996	URL_Start	512	0
52	201880	LogData	0	Vurderingsproblem
53	219749	LeftMouseButton	1	lybe tenker tur lengde frem og tilbake
54	263786	LogData	0	Vurderingsproblem
55	964481	LogData	0	lybe problem med present
56	268884	LeftMouseButton	1	lybe sÅrre fonter i tekst
57	277086	URL_End	1024	0
58	278327	URL_Start	512	0
59	287749	LeftMouseButton	1	253
60	293988	LeftMouseButton	1	271
61	294870	LeftMouseButton	1	273
62	297951	LeftMouseButton	1	276
63	298213	LeftMouseButton	1	264

Oppsummering

Innenfor kvalitetsarbeidet i statistikkbyråene er det viktig å forsøke å måle kvaliteten i de ulike produksjonsprosessene. Generelt er det problematisk å knytte prosessdata til resultatdata. Den ”eksterne” prosessen å samle inn data er spesielt viktig å måle siden den er helt avgjørende også for kvaliteten på sluttproduktet. Etter elektronifiseringen av datafangsten de siste 10 – 15 år har vi fått nye datakilder som kan hjelpe oss til å måle denne prosessen. Paradata kan sammen med teori og kvalitative metoder bidra til å kaste lys over spørsmål og svar prosessen i datafangst. Eytracker teknologi utnytter både tradisjonell kognitiv testing og paradata, sammen gir dette et sterkt måleinstrument for å avdekke målefeil og forbedre våre skjemaer i planleggingsprosessen. Når datainnsamlingen foregår ved hjelp av elektroniske selvutfyllingskjema kan vi selvsagt ikke samle inn blikksporingsdata. Derimot kan vi telle antall handlinger (klikk) og tiden fra en handling til den neste, antall endringer i skjemaene, aktiviserte hjelpefunksjoner og aktiviserte feilmeldinger. Vi kan tenke oss at et velfungerende skjema vil være et skjema som krever relativt få klikk, som ikke genererer feilmeldinger og som respondenten fyller ut relativt raskt.

Litteratur

Haraldsen, Gustav, Øyvin Kleven og Anne Sundvoll (2005) "Big Scale Observations gathered with the help of Client Side Paradata". Paper presented at the Workshop on Methodology of Usability and Questionnaire of CASI Questionnaires, Heerlen, The Netherlands April 19-21.

Haraldsen, Gustav, Øyvin Kleven and Margareta Stålnacke (2006) "Paradata indications of problems in Web Surveys". Paper presented at The European Conference on Quality in Survey Statistics, Cardiff, Wales, April 24-26.

Haraldsen, Gustav, Øyvin Kleven and Anne Sundvoll (2010) "Quality Indicators in Data Collection". Paper presented at The European Conference on Quality in Official Statistics, Helsinki, Finland, May 3-6.

Heerwegh, Dirk. 2002. "Describing response behavior in Web surveys using client-side paradata". Paper presented at the International Workshop on Web Surveys, Mannheim, Germany, October 17-19.

———. 2003. "Explaining response latency and changing answers using client-side paradata from a Web survey". In *Social Science Computer Review* 21 (3): 360-73.

Køber, Tonje, Nina Arnesen, Matz Ivan Faldmo, Jenny Linnerud, Peder Næs, Grete Olsen, Bente Thomassen, Harald Utne og Li-Chun Zhang (2008) *Virksomhetsmodell for Statistisk sentralbyrå*. Oslo-Kongsvinger. Notater 2008/47

Snijkers, Ger and Mattijn Morren (2010) "Improving web and electronic questionnaires: The Case of audit trails. Paper presented at The European Conference on Quality in Official Statistics - Helsinki. Finland 3-6 May.

Stern, Michael J. (2008) "The use of Client-side Paradata in Analyzing the Effects of Visual Layout on Changing Responses in Web Surveys" in *Field Methods*, Vol. 20, No. 4, November 2008, Sage Publications

Sudman, S., N. Bradburn, and N. Schwarz. 1996. *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

UNECE (2009) Generic Statistical Business Process Modell. Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS). Prepared by the UNECE Secretariat.