

Notat 8/03

Om datakvalitet i bredbåndskartlegging

Om datakvalitet i bredbåndskartlegging

Innhold:

1	INNLEDNING	1
2	METODE.....	1
3	KVALITETSSIKRING	1
3.1	Punching.....	1
3.2	Etterkontroll	1
4	KONTROLLUNDERSØKELSE	2
4.1	Svarfordeling.....	2
4.2	Enkeltobservasjoner	3
5	KONKLUSJONENE STÅR FAST	5

1 Innledning

Computerworld har tirsdag 29. januar 2003 et oppslag om ECON og Teleplans bredbåndsrapport. Det fremkommer påstander om tastefeil og tallrot. Computerworld har kontaktet 19 kommuner, og funnet 12 feil i et tabellvedlegg som har fulgt med rapporten.

2 Metode

For undersøkelsens formål var det viktig å få svar fra flest mulig kommuner. Vi vurderte derfor at en utvalgsundersøkelse hvor bare et fåtall kommuner ble kontaktet ikke var aktuelt. Med en slik innfallsvinkel ville det vært mulig med en telefonundersøkelse og følgelig større kontroll med datakvaliteten. Grunnen til at vi valgte å kontakte alle kommunene, var at en sentral del av mandatet var å gi et overslag for hva det vil koste å gi alle kommuner et bredbåndstilbud til konkurransedyktige priser. Uten å kjenne til "demografien" i kommunene mht. antall, størrelse og lokalisering av institusjonene ville det være vanskelig å lage et utvalg som vi med rimelig trygghet kunne anta var representativt.

Vi valgte å la skjemaet gå til rådmennene med et følgebrev fra Statsråden. På denne måten fikk undersøkelsen tyngde og vi antok at rådmennene selv ville sørge for at riktige personer fikk ansvar for å fylle ut skjemaene. Vi fant at dette var en bedre fremgangsmåte, for eksempel sammenlignet med en henvendelse direkte til kommunenes IT-sjefer.

Denne prosedyren og selve spørsmålsstillingen i skjemaene ble utviklet i samråd med kontaktpersoner i utvalgte "pilot"-kommuner.

3 Kvalitetssikring

3.1 Punching

To feilene som Computerworld har funnet er feil innlegging av data (punching) av skjemaene. Vi har valgt ut ca 150 skjemaer tilfeldig. Dette utgjør 6 prosent av alle skjemaene. Hvert skjema inneholder 5 data, det vil si at til sammen 750 observasjoner er kontrollert. Kontrollen har skjedd ved at en person har lest opp svarene, mens undertegnede har sjekket regnearket. Av disse har vi funnet 6 feil, det vil en feilprosent på 0,67 prosent. Ingen av disse påvirker konklusjonene fra rapporten. Dette gir oss grunn til å regne med at kvaliteten er tilstrekkelig høy på dette punktet.

3.2 Etterkontroll

Teleplan har gått gjennom alle data for å luke urimelige svar. Metoden har vært en antagelse om at aksesteknologien har vært mer pålitelig informasjon enn

kapasitet. De urimelige svarene er derfor tallene for kapasitet skjønnsmessige korrigert. Denne korrigerte listen er imidlertid ikke tatt inn i vedlegget.

ECON har brukt ca ¼ månedsverk til å følge opp spørreundersøkelsen.

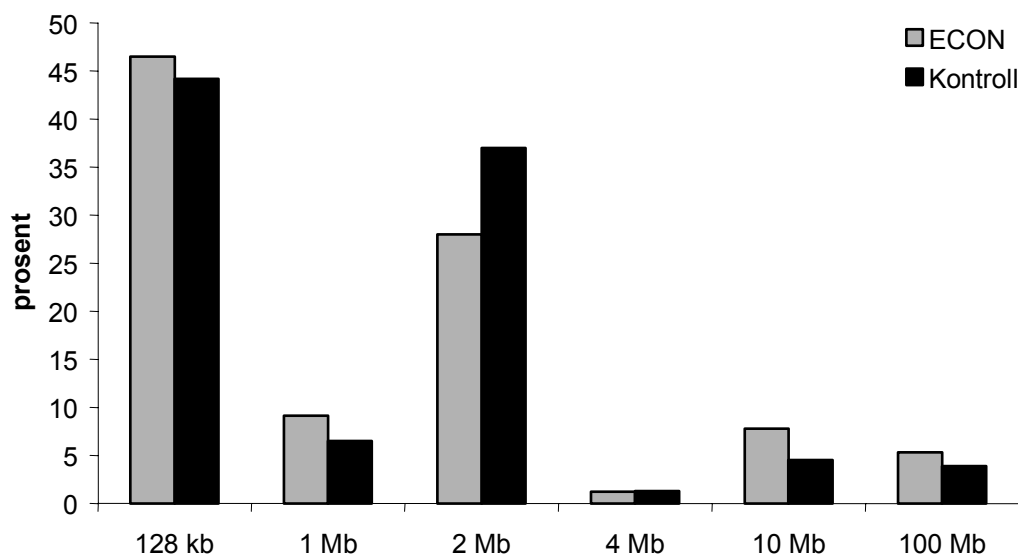
- Ca 60 kommuner har på eget initiativ ringt til ECON for å oppklare misforståelser, uklarheter i skjema osv. Disse er da blitt muntlig veiledet.
- Vi har hatt en person på fulltid for å ringe rundt til kommunene. Dette var primært for å øke svarprosenten. Vedkommende har så langt som mulig forsøkt å identifisere personen i kommunen som har hatt ansvaret for å fylle ut skjemaene. Vi antar derfor å i betydelig grad ha oppdaget om skjemaer er blitt sendt til feil person. I tilfeller hvor det var åpenbare feil i skjemaene har vi også ringt tilbake for å sjekke. Dette har f.eks. vært feil av typen at en kommune oppgir å ha 60 lokasjoner.

4 Kontrollundersøkelse

4.1 Svarfordeling

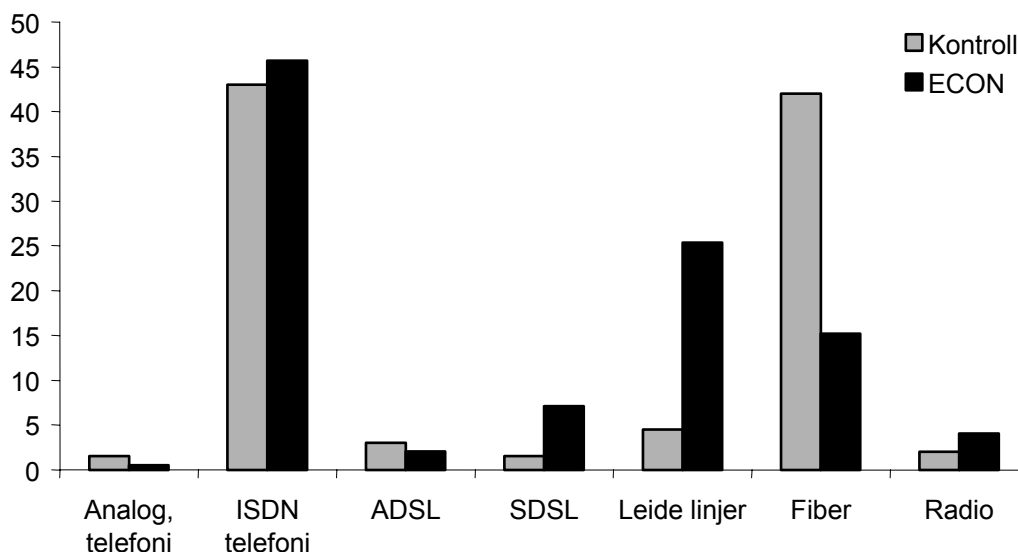
For å foreta en ekstra kontroll av datamaterialet, har ECON engasjert Analysehuset for å hente inn data for 200 kommuneinstitusjoner på nytt gjennom telefonintervjuer. De 200 kommunene er valgt av Analysehuset. Disse dataene er sammenholdt med opplysningene som tidligere har vært samlet inn fra de samme institusjonene. Figuren viser at *dataene er nærmest overlappende*. Det er ingen systematiske avvik som tilsier at de konklusjonene som er trukket i rapporten skal være feilaktige.

Figur 4.1 *Kapasitet for Internettilknytning i norske kommuner, ECONs opprinnelige tall og kontrollundersøkelsen*



Kilde: ECON og Analysehuset

Figur 4.2 Internettilknytning i norske kommuner etter aksesteknologi, ECONs opprinnelige tall og kontrollundersøkelsen



Kilde: ECON og Analysehuset

Figuren viser at det er stor grad av overensstemmelse i fordelingen mellom telefoni og bredbåndsteknologier mellom ECONs tall og tall fra kontrollundersøkelsen. Den viktigste forskjellen i de to undersøkelsene er fordelingen mellom SDSL, Leide linjer og fiber. Det må understrekes at leide linjer ikke er en egen teknologi, men en kontraktsform mellom bruker og Internettleverandør. I den store andelen Leide linjer i ECON-tallene vil teknologien kunne være både xDSL, fiber og radioforbindelse.

4.2 Enkeltobservasjoner

Når vi sammenligner ECONs data, kontrollundersøkelsen og dataene som Computerworld har innhentet er det betydelige avvik for enkeltkommuner. Sarpsborg kommune har oppgitt en kapasitet på 620 Mbps til både Computerworld og ECON, men 2 Mbps til Analysehuset. En annen kommune har oppgitt 100 Mb til ECON, 2 Mb til Analysehuset og 128 kbps til Computerworld. En tredje kommune oppgir 2 Mbps til Computerworld, 128 kbps til Analysehuset og 100 Mbps til ECON.

Så store avvik er imidlertid unntak. Det er imidlertid svært få (ca ¼) som gir *eksakt* samme svar i ECONs undersøkelse som i kontrollundersøkelsen. Langt flere avgir svar som er rimelig like, f.eks. at enten hastighet eller teknologi er sammenfallende i de to undersøkelsene.

For enkeltobservasjoner vil det altså være feilkilder både ved den opprinnelige datainnsamlingen og ved telefonintervjuene. En slik feilkilde i den opprinnelige undersøkelsen er at noen respondenter oppgav kapasiteten i det interne datanettverket og ikke for Internettilknytningen, slik det er spurt om. En feilkilde ved å sammenligne tallene fra Analysehuset og ECON er at tallene er hentet inn med et halvt års mellomrom. Institusjonene kan ha endret tilknytningsform og ulike personer kan ha besvart spørsmålene. Forskjellig i metode gir også feilkilder. En telefonundersøkelse gir respondentene mindre tid til å svare sammenlignet med et

skriftlig skjema. I telefonintervjuene er det mange IT-ansvarlige på skolene som rett og slett ikke vet Internetkapasiteten. I ECONs undersøkelse er denne feilkilden forsøkt unngått ved at skjemaene er sendt sentralt i kommunen slik at man har kunnet koordinere innsamlingen av data ut i fra om IT-ansvaret er sentralisert eller ikke. På den annen side gir telefonintervjuer bedre muligheter til å presisere problemstillingen og følgelig er sannsynligheten for misforståelser mindre.

Tabell 4.1 Sammenligning av svar for enkeltinstitusjoner

(1) Eksakt likt svar	23 %
(2) Eksakt lik hastighet	41 %
(3) Eksakt lik teknologi	51 %
(4) Enten eksakt lik hastighet eller lik teknologi	67 %
(5) Lik hastighets og teknologiklasse	55 %
(6) Lik hastighetsklasse	66 %
(7) Lik teknologiklasse	72 %
(8) Enten lik hastighets- eller teknologiklasse	77 %
(9) Som 5, men med høyere hastighet	71 %
(9) Som 8, men med høyere hastighet	88 %

Kilde: ECON og Analysehuset Tallene i tabellen illustrerer altså at feilkildene for enkeltobservasjoner er betydelige. Under halvparten hav kommunene har gitt eksakt det samme svaret i ECONs og Analysehusets undersøkelse.

Dersom vi i stedet grupperer dataene i klasser slik at vi sammenligner svarene for hastigheter i samme kategorier som i figur 4.1 og svarene for teknologi i kategoriene

1. Telefoni
2. DSL
3. Leide linjer og fiber
4. Radio

blir svarene mer sammenfallende, et sted mellom 55 og 72 prosent.

Hvis vi videre sammenligner etter om kommunene gir det likt svar mht. *enten teknologi eller hastighet* og tar hensyn til at kommunene kan ha oppgradert Internettilknytningen mellom september 2002 og februar 2003 får vi et sammenfall på mellom 77 og 88 prosent.

Vår konklusjon er at det er krevende å innhente helt presise tall for kommunenes bredbåndstilknytning. Kontrollundersøkelsen indikerer imidlertid ikke at dataene i den opprinnelige undersøkelsen fortjener karakteristikken ”skivebom”.

5 Konklusjonene står fast

Hovedkonklusjonene i Bredbåndskartleggingen var at over halvparten av kommunene er tilknyttet bredbånd i dag, og det vil koste mellom 0,8 og 1,7 mrd NOK å oppnå målene i eNorge planen. Så lenge det ikke er *systematiske avvik* vil ikke feil i enkeltobservasjoner røkke disse konklusjonene. Analysehuset har altså ikke funnet slike systematiske avvik. Både i ECONs tallmateriale og i kontrollundersøkelsen er andelen ISDN-basert Internettilknytning rundt 45 prosent. Videre viser begge undersøkelsene at rundt en tredel av kommunene har en kapasitet på rundt 2 Mbps. Det er derfor ikke grunnlag for å anta at vår kartlegging under vurderer hvor mange kommuner som vil trenge støtte for å få bredbånd.

En illustrasjon på robustheten i kostnadsberegningen er at kun 2 av de 12 ”feilene” som Computerworld har funnet innebærer endrede forutsetninger. Årsaken er todelt: i beregningsgrunnlaget inngår kun institusjoner som ikke allerede har etablert bredbåndstilknytning og som ligger i områder hvor vi ikke regner med vil få et kommersielt bredbåndstilbud innen 2005. Videre er det likegyldig for beregningene om en av aktuelle institusjonene har 128 eller 256 kbps i dag.