

Sigbjørn Strømsnes og Thomas Hagen

**Datafangst lønnsstatistikk
– en systematisk gjennomgang
av prosesser fra utsending av
skjema til data er ferdig for
tabellproduksjon**

Innholdsfortegnelse

1 Innledning	3
2 Bakgrunn og avgrensning	4
3 Generelt om den kvartalsvise lønnsindeksen	4
4 Sentrale prosesser	5
4.1 Utsendelse	5
4.2 Mottak	5
4.2.1 Telefontjenesten.....	5
4.2.2 Mailboks	6
4.3 Dataregistrering og innlesing av rapporterte data	6
4.3.1 Manuell registrering.....	6
4.3.2 Optisk lesing	6
4.3.3 Elektronisk innrapportering	6
4.3.4 Innlesing til tabell	7
5 Forbedringsmuligheter	7
5.1 Raskere innrapportering av data.....	8
5.2 Raskere bearbeiding av innkomne data.....	9
6 Konklusjon	16

Figurer

<i>Figur1: Foretakenes innkvitteringsdato</i>	8
<i>Figur2: Antall foretak fordelt etter innrapporteringsform</i>	9
<i>Figur3: Antall sysselsatte personer fordelt etter innrapporteringsform</i>	10
<i>Figur4: Antall foretak fordelt etter antall dager fra innkvitteringsdato til foretaket er ferdig innlest på tabell</i>	11
<i>Figur5: Antall dager fra et foretak er innkvittert til det er innlest på tabell etter innrapporteringsform</i>	12
<i>Figur6: Antall foretak rapportert etter kravspesifikasjon fordelt etter antall dager fra innkvitteringsdato til de er innlest på tabell</i>	12
<i>Figur7: Antall foretak som er optisk lest fordelt etter antall dager fra innkvitteringsdato til de er ferdig innlest på tabell</i>	13
<i>Figur 8: Antall foretak manuelt registrert fordelt etter antall dager fra innkvitteringsdato til de er ferdig innlest på tabell</i>	14

Vedlegg

<i>Flyttdiagram for produksjonsprosessen</i>	17
<i>Formålsbeskrivelse</i>	18

1 Innledning

Notatet er sluttrapporten fra prosjektet Datafangst lønnsstatistikk med utgangspunkt i kvartalsstatistikk lønn. Arbeidet med prosjektet har inngått som et av prosjektene i Statistisk sentralbyrås satsing på systematisk kvalitetsarbeid (SKA).

Prosjektgruppen hadde følgende sammensetning:

- Prosjektleder Sigbjørn Strømsnes (Seksjon for inntekts- og lønnsstatistikk)
- Kvalitetslos Ole Rognstad (Seksjon for primærnæringer)
- Prosjektmedarbeider Thomas Hagen (Seksjon for inntekts- og lønnsstatistikk)
- Prosjektmedarbeider Bjørg Ekørhol (Seksjon for inntekts- og lønnsstatistikk)
- Prosjektmedarbeider Aud Kari J. Holt (Seksjon for datafangst)
- Prosjektmedarbeider Åse Krogsrud (Seksjon for datafangst)
- Prosjektmedarbeider Sindre Børke (Seksjon for datafangst)

Styringsgruppe for prosjektet:

- Sindre Børke (Seksjon for datafangst)
- Per Ove Smogeli (Seksjon for inntekts- og lønnsstatistikk)

2 Bakgrunn og avgrensning

Statistisk sentralbyrå (SSB) produserer årlig lønnsstatistikk for ulike næringer som skal belyse lønnsnivå, strukturer og endringer i lønn. I tillegg produseres hvert kvartal en indeks som skal belyse endring i lønn over året.

Med bakgrunn i økte krav fra brukerne til aktualitet i den kvartalsvise lønnsindeksen, fra 90 dager til 60 dager etter tellingstidspunktet, ønsker seksjon 420 en gjennomgang av produksjonsprosessene i datafangsten til lønnsstatistikken. Brukerne av den kvartalsvise lønnsindeksen er Eurostat, IMF, Det tekniske beregningsutvalget for inntektsoppgjørene og Nasjonalregnskapet. I tillegg til å være en selvstendig statistikk som belyser lønnsendringer over året er lønnsindeksen også en viktig informasjonskilde til byggekostnadsindeksen, lastebilkostnadsindeksen og konsumprisindeksen.

Målsettingen med prosjektet er å belyse alle prosesser i forbindelse med datafangsten og vurdere om noe kan endres for å øke aktualiteten, se vedlagt formålsbeskrivelse. Vi vil benytte den kvartalsvise lønnsindeksen som eksempel i dette prosjektet, men eventuelle resultater fra prosjektet vil være relevant for alle lønnsstatistikkenes.

Den kvartalsvise lønnsindeksen omfatter mange prosesser, fra trekking av utvalg, tilrettelegging av skjema, utsending til oppgavegiver, innsamling av data, kontroll av data, innlasting på tabell, produksjon av tabeller og presentasjon (se vedlegg 1). Prosjektet fokuserer kun på prosessene fra og med utsending av skjema til oppgavegiver og fram til data er tilrettelagt på tabell og klare til å benyttes i tabellproduksjon. Dette vil si at vi ikke vil gå inn på de første og siste prosessene i arbeidet med den kvartalsvise lønnsindeksen. Årsaken til dette er at vi ønsker å avgrense prosjektet i størst mulig grad mot det som er målsettingen, nemlig å redusere produksjonstiden i tråd med internasjonale og interne krav.

3 Generelt om den kvartalsvise lønnsindeksen

Den kvartalsvise lønnsindeksen omfatter i år 2000 om lag 1 400 foretak fordelt på 6 forskjellige næringshovedområder:

- Olje- og gassutvinning og bergverksdrift
- Industri
- Bygge- og anleggsvirksomhet
- Varehandel
- Samferdsel
- Forretningsmessig tjenesteyting og eiendomsdrift

Målsettingen med indeksen er å belyse endring i lønn over kvartalet. Det er med andre ord ingen målsetting å skulle vise ulike lønnsnivåer hvert kvartal, noe som ville krevd et større og mer omfattende utvalg. Indeksen presenteres for to forskjellige lønnsbegreper. Det ene omfatter kun endringen i utbetalt avtalt lønn ved utgangen av kvartalet. Det andre lønnsbegrepet omfatter endringen i månedsførtjenesten, noe som i tillegg til utbetalt avtalt lønn også inkluderer uregelmessige tillegg og eventuelle bonusutbetalinger.

Hvert 1. kvartal trekkes et stratifisert utvalg basert på foretakenes størrelse og opplysninger fra den årlige lønnsundersøkelsen hvert 3. kvartal, og samme utvalg benyttes et helt år. Skjema til

lønnsindeksen sendes til foretak som rapporterer per individ fordelt på bedrifter. Skjemaet inneholder følgende variabler:

- Fødselsnummer
- Avtalt arbeidstid per uke
- Lønnsperiode (om man har timelønn, 14-dagers lønn, månedslønn, provisjonslønn)
- Utbetalt avtalt lønn i henhold til lønnsperiode
- Uregelmessige tillegg hittil i år
- Bonuser, provisjoner mv. hittil i år
- Overtidsutbetaling hittil i år
- Overtidstimer hittil i år

SSB ønsker at flest mulig skal rapportere statistikken elektronisk og har derfor utarbeidet en egen kravspesifikasjon for innrapportering av lønnsstatistikk som mange har tatt i bruk. Første gang statistikken skal leveres på denne måten kreves det en tilpasningsjobb, men etter at dette er gjort skal det i prinsippet ikke være vanskeligere for foretaket enn å sette inn en diskett, angi tidsperiode og "trykke på knappen". Det er i tillegg utarbeidet en kravspesifikasjon for innrapportering på regneark. Av de som ikke leverer elektronisk rapporterer de fleste på skjema, men det er også noen foretak som sender inn datalister.

4 Sentrale prosesser

4.1 Utsendelse

Skjemaene til den kvartalsvise lønnsindeksen blir sendt til foretakets postadresse. Hver bedrift i foretaket får sitt skjema og det blir sendt ut en rettledning og ett følgebrev til hvert foretak. Vi skiller bedrifter etter rapporteringsmåte, dvs. om de rapporterer elektronisk eller manuelt. Bedrifter som ved den siste årlige lønnsstatistikken rapporterte manuelt får tilsendt tilleggsark etter antall sysselsatte. De som ved forrige innsending leverte på diskett skal ikke ha tilleggsark, som følge av at de ikke benytter skjemaet i rapportering (med unntak av forsiden, som uansett skal returneres).

4.2 Mottak

Få dager etter utsendelsen begynner skjemaene å komme inn. Det rulleres på hvem som går gjennom posten og mange skjema må legges til side for å ringes på. Det som vanligvis skjer er at bunken med foretak som må ringes vokser og skjemaene blir "gamle" før vi får tatt dem. Det er ikke alltid vi treffer kontaktpersonen første gangen vi ringer og vi bruker mye tid på noen av skjemaene. Dette er lite hensiktsmessig når vi tenker på at dette er utsendinger som kan resultere i tvangsmulkt og det krever orden i korrespondansen.

Mottaksrutinen er svært viktig for datafangsten, da vi i oppslagsbildet i delregisteret finner all historikk til foretaket. Dette er nyttig både når vi får telefoner fra oppgavegiver og for vår egen del i behandlingen av foretaket. Merknadsfeltet må benyttes flittig. Dette er kanskje den delen av datafangsten som krever mest nøyaktighet og det er derfor viktig at det utøves ekstra påpasselighet i denne prosessen og at vi tar oss tid til å få alt på plass.

4.2.1 Telefontjenesten

Det er opprettet en telefonslynge hvor 8 personer har vært involvert. I perioder like etter utsending, rundt fristens utløp og etter purrerunder er pågangen størst.

Mange av telefonhenvendelsene er fra oppgavegivere som har kastet eller forlagt sitt skjema. I løpet av tellingen skrives det ut mange nye skjema for hånd, noe som er svært tidkrevende. Det bør etableres en rutine for preprinting av slike skjema.

4.2.2 Mailboks

Høsten 2000 ble det opprettet en mailboks med adresse `lonnstat@ssb.no`, en adresse som er oppgitt i rettledningen. Mailboksen besvares daglig og har vært betjent av flere personer som har fordelt ansvaret (turnusliste). Søknader om utsettelse blir videresendt til fagkontoret, mens de øvrige henvendelsene blir behandlet på seksjon 450. Alle henvendelser blir skrevet ut og samlet i ulike omslag. Alle mailer som blir besvart av seksjon 450 påføres organisasjonsnummer og næringskode og det noteres i merknadsfeltet hva som er gjort. Når mailen er ferdigbehandlet slettes den i mailboksen. Konklusjonen er at erfaringene er gode og at mailboksen har kommet for å bli. Den avlaster telefonen og man unngår det samme stressmomentet som det er å sitte med oppgavegiver på telefon. Det blir bedre tid til å vurdere henvendelsen og betjene spørsmålene som kommer.

4.3 Dataregistrering og innlesing av rapporterte data

4.3.1 Manuell registrering

Det har generelt vært lite problemer med manuell registrering, men noen foretak krever imidlertid store ressurser. Noen oppgavegivere lager sine egne lister, gjerne to stykker, som skal kobles. Alt som kommer til manuell registrering må punches av SSB, og det er derfor mye jobb som må gjøres med disse foretakene.

4.3.2 Optisk lesing

I utgangspunktet skal skjema for lønnsstatistikk leses optisk. Ettersom vi har fått mer og mer erfaring med optisk lesing er det ingen tvil om at kvaliteten er varierende. Det er mulig at skjema for lønnsstatistikk ikke er spesielt godt egnet til optisk lesing. På skjemaet er det små skille tegn mellom tallene som ofte blir lest som tall. Det er et problem at tallkolonnene på skjema har lite luft mellom seg, noe som skaper problemer for lesingen. Det er stort sett ikke noe system på hva som går galt med lesing av tall, men det vanligste problemet er at "ett-tall med hake" skaper problemer. Noen godkjennes, andre ikke. De fleste av disse blir lest som sju-tall, noe det ofte kan være veldig vanskelig å avsløre. Når det gjelder avlønningsperiode har vi 7 (provisjonslønte) som gyldig kode. Disse må godkjennes manuelt av den som verifiserer på grunn av problemene med ett-tall og sju-tall. Tre-tall som går nedenfor linja blir ofte til to-tall.

4.3.3 Elektronisk innrapportering

Stadig flere foretak innrapporterer elektronisk og kvaliteten på det som kommer inn er varierende. Det sendes blant annet inn disketter som er ødelagte eller har feil som gjør dem ubrukelige. Dette medfører at SSB må etterlyse disse diskettene på nytt. Dersom en diskett tar for mye tid å redigere kan en mulighet være å skrive den ut og foreta manuell registrering. Dette er alltid en avveining mot hvor mange ansatte dette gjelder og graden av problemer med disketten.

En god del oppgavegivere sender statistikken på regneark. I rettledningen står det at SSB skal kontaktes dersom foretaket ønsker denne rapporteringsformen, men det blir ofte ikke gjort. Disse kan være problematiske.

En gjennomgangsfeil for de som leverer etter kravspesifikasjon for elektronisk innberetning er at siste lønnsutbetaling ikke kommer med. Det virker som om de ansvarlige som kjører lønn hos foretakene ikke helt vet hvordan de skal kjøre ut rapporten. Vi er i kontinuerlig dialog med oppgavegiverne om dette og rapporteringen er gradvis blitt bedre, selv om det fortsatt ikke er bra nok.

4.3.4 Innlesing til tabell

Etter at innkomne data er blitt sjekket visuelt og i en automatisk diskettsjekk med hensyn på at nødvendige data er med så overføres filene til Unix, på kataloger klargjort for innlesing til tabell.

Selve innlesingen er relativt teknisk og det er kontor 403 som står for rutinene omkring dette, men oppsummert går den ut på at innlesingen av data er bygget opp rundt et eget program. Her knyttes posisjoner i de ulike filene som er lagt til rette på forhånd mot spesifikke kolonner i tabellen. Filene blir så kopiert til ulike områder på unix, samt at under innlesingen produserer programmet nye logfiler og feilfiler. Det er seksjon 450 som kjører innlesing av data til oracle-tabellen, og som kontrollerer loggen etter hver innlesing. Loggen som blir skrevet ut etter hver kjøring gir en oversikt over hva som havner i tabellen, og hva som havner i feiltabellen.

Ikke alle disketter som ser riktige ut lar seg lese inn på tabell. Sammen med kontor 403 er det brukt mye tid for å finne ut hva som har gått galt. Noen disketter er overført opptil 4 ganger før vi har fått dem på tabell. Kontor 403 har også benyttet manuell innlesing på en del disketter. Alt dette er svært tidkrevende.

Data i feilfilene havner i egen feiltabell. Denne tabellen jobber seksjon 420 seg gjennom med jevne mellomrom, rydder opp i og retter det som gjør at dataene ikke kan leses rett i tabellen. Deretter blir disse dataene overført til tabellen.

Dataene som havner i logfil settes sammen til resultatinformasjon som automatisk blir sendt på e-post til saksbehandlere på seksjon 420 og 450.

For å sjekke om noen innkomne data ikke er blitt lest inn verken på tabellen eller feiltabellen kjører seksjon 420 et sas-program som kontrollerer om vi finner igjen data fra alle de innkvitterte bedriftene på en av tabellene. Finner vi ikke *noen* observasjoner fra en bedrift sjekkes det om denne bedriften er blitt lagt ut på unix og blitt forsøkt lest inn, eller om bedriften er til behandling (kontroll eller oppretting) hos seksjon 450. Dersom den er blitt forsøkt lest inn, må den kontrolleres på nytt og deretter legges en gang til på innlesingskatalog i unix. Denne kontrollen med å kjøre et slikt sas-program har vi per i dag ikke en fast rutine for, men det bør definitivt opprettes. Samtidig vil det være nyttig hvis en slik test også kunne utvides til å finne enkeltobservasjoner (fra bedrifter der andre observasjoner faktisk er blitt lest inn) som eventuelt ikke havner i noen tabell.

Ved innlesingen foretas det noen automatiske kontroller av dataene. Disse går i hovedsak ut på å sjekke at verdiene på enkelte variabler er innenfor gitte grenser eller at gyldige koder er oppgitt. Per i dag er det kontroll på organisasjonsnumrene, avtalt arbeidstid, kode for lønnsperiode, utbetalt avtalt lønn, uregelmessige tillegg, bonus, overtidsgodtgjørelse og overtidstimer. Disse kontrollene tilsier ikke om dataene er korrekte, bare at de er innenfor gitte grenseverdier.

5 Forbedringsmuligheter

Etter å ha foretatt en gjennomgang av sentrale prosesser i datafangsten kan man skille datafangsten i to sentrale prosesser som begge er viktige for å få en raskere produksjonstid:

- Raskere innrapportering av data
- Raskere bearbeiding av innkomne data

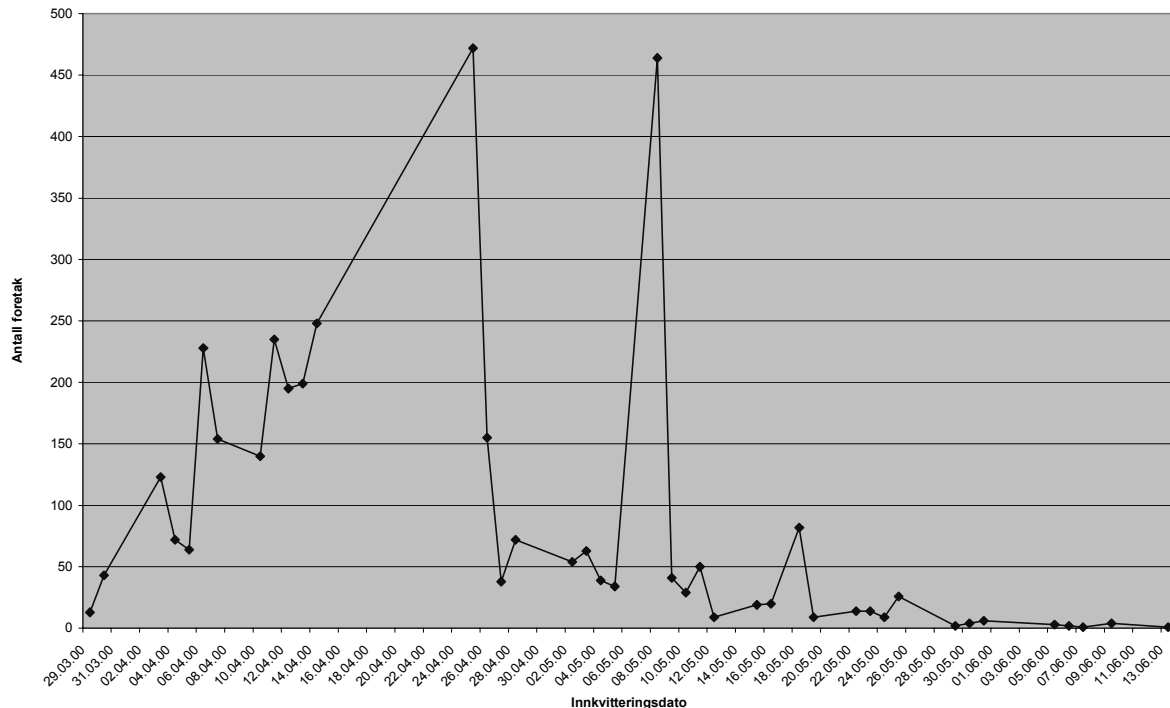
Disse to prosessene er de viktigste når man skal ha som målsetting å gjennomføre en hurtigere og mer strømlinjeformet datafangst.

5.1 Raskere innrapportering av data

Denne delen av datafangsten er utvilsomt den minst sammensatte, selv om det ikke nødvendigvis er noen enkel prosess å skulle gjennomføre en forbedring.

Figuren nedenfor viser når dataene til 1. kvartal 2000 ble innkvittert:

Figur 1: Foretakenes innkvitteringsdato



Figur 1 viser at foretak blir innkvittert kontinuerlig, helt fra kvartalets utløp og fram til midten av juni som her ble satt som siste frist for innlesing på tabell.

Følgende tiltak kan gi raskere innrapportering av data:

- Seksjon 420 må være ute i god tid når det gjelder trekking av utvalg og å få sendt skjema til preprinting på Kopsisenteret. Det er en fordel at alt kan være pakket ferdig i god tid før utsending. Stress kan medføre at det gjøres feil i pakkingen, noe som igjen kan få konsekvenser senere i datafangsten, både med feil innrapportering og mange unødvendige telefoner.
- Svarfristen må ikke være for lang. Tidligere har denne vært på om lag 14 dager etter kvartalets utløp. Denne fristen kan vi med fordel kutte ned til 1 uke, men dette forutsetter at utsendingen av skjema blir foretatt i god tid før kvartalets utløp (om lag den 20. i måneden før).
- Utsending av brev med "varsel om tvangsmulkt" og "vedtak om tvangsmulkt" må skje etter en stram tidsplan. Korttidsstatistikk har ikke de samme krav til lengde på frister som årsstatistikken har, men foretaket bør uansett få minst 1 ukes tidsfrist både når det mottar brev med varsel om tvangsmulkt og vedtak om tvangsmulkt.
- Mange foretak søker om utsettelse av innlevering, noe som selvfølgelig skal være tillatt. Det kan imidlertid virke som om noen helt bevisst spiller på dette for å kjøpe seg tid. Det er derfor viktig at det settes opp korte utsettelser og at alle søknader om dette må være skriftlig. Seksjon 420 skal ta stilling til om et foretak får utsettelse og hvor lang utsettelsen eventuelt skal være.
- Store foretak må følges opp på et tidlig tidspunkt. Noen foretak er det mer essensielt å få inn enn andre, og det er viktig at disse blir oppringt dersom de ikke leverer. Kanskje kan det være internt

rot som gjør at foretaket ikke leverer, eller at adressen som er benyttet ikke har vært korrekt. Generelt er det derfor viktig å få en personlig kontakt med kontaktpersonene på de "store lokomotivene".

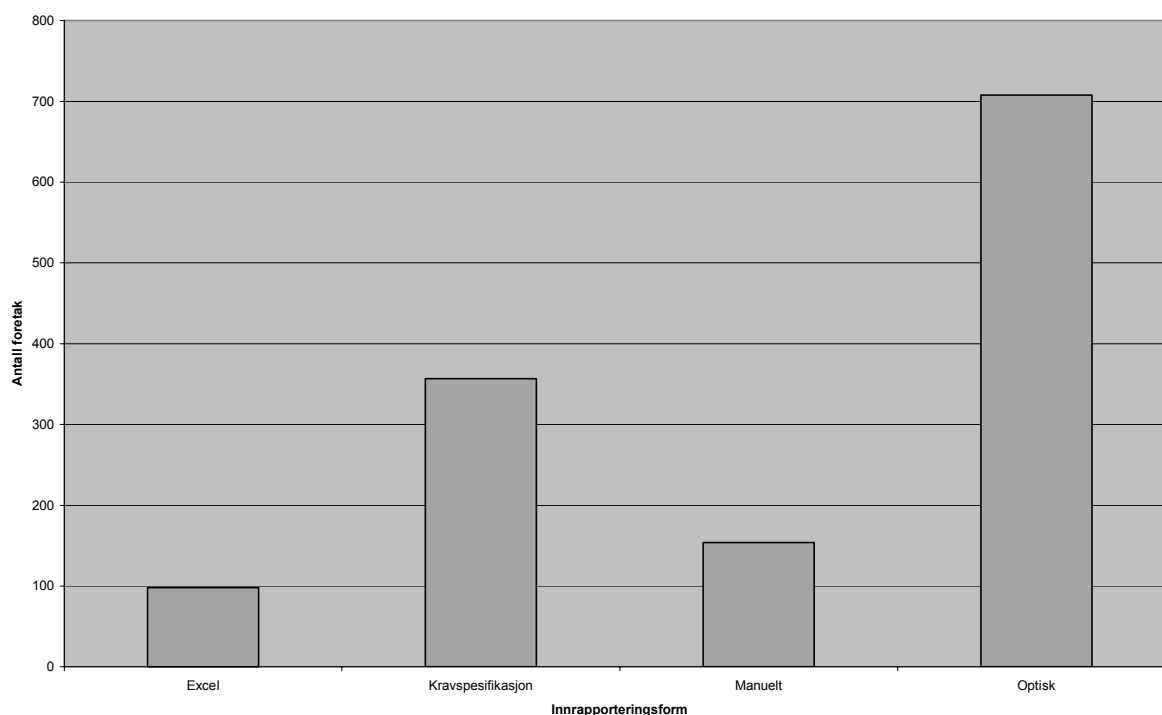
Dersom punktene ovenfor blir gjennomført konsekvent bør vi ha en langt større andel inne på et tidligere tidspunkt enn tidligere, noe som er helt avgjørende for om vi skal greie å forbedre produksjonstiden.

5.2 Raskere bearbeiding av innkomne data

Det er begrenset hvor mye man kan spare inn av tid ved å sette kortere frister i datafangsten og ved å sende ut skjema tidligere. Derimot kan det være mye tid å spare på å få en bedre og mer strømlinjet prosess fra dataene kommer i hus og fram til de ligger ferdig på tabell, klar til å benyttes i statistikkproduksjonen.

Behandlingen av dataene er fundamentalt forskjellig ut i fra hvilken innrapporteringsform de er kommet inn på. Som det er beskrevet tidligere i notatet er det konkrete metoder som benyttes i bearbeidingen på hver enkelt rapporteringsform. Viktig er det allikevel å se på hvor mye som rapporteres på den enkelte innrapporteringsform og hvor mye tid som benyttes på hver av disse.

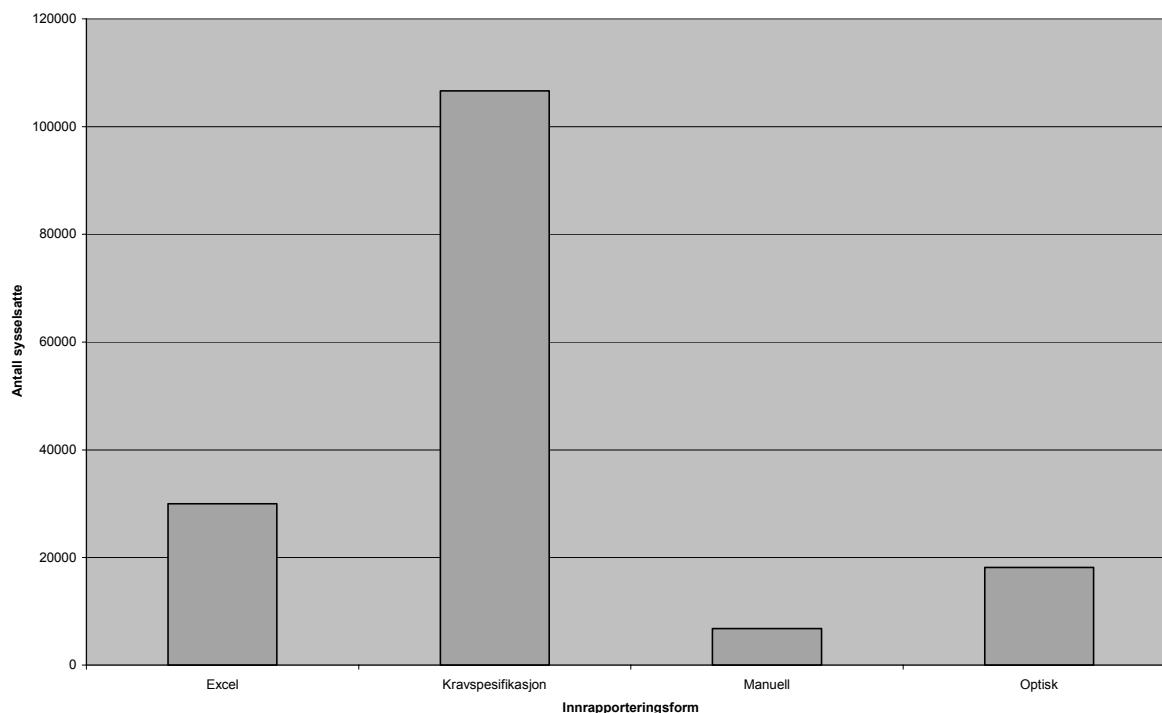
Figur 2: Antall foretak fordelt etter innrapporteringsform



I figur 2 er det gitt en oversikt over hvor mange foretak som rapporterer på de ulike innrapporteringsformene. Som man ser av figuren er det en klar overvekt av foretak som rapporterer på skjema og som dermed blir lest optisk. Det er for øvrig mer enn 350 foretak, eller om lag 25 prosent, som rapporterer ifølge SSBs kravspesifikasjon.

Utvalget til den kvartalsvise lønnsindeksen består av både store og små foretak. Størrelsen på foretakene er så forskjellig at det er enda viktigere å se på sysselsettingsfordelingen på de forskjellige innrapporteringsformene.

Figur 3: Antall sysselsatte personer fordelt etter innrapporteringsform

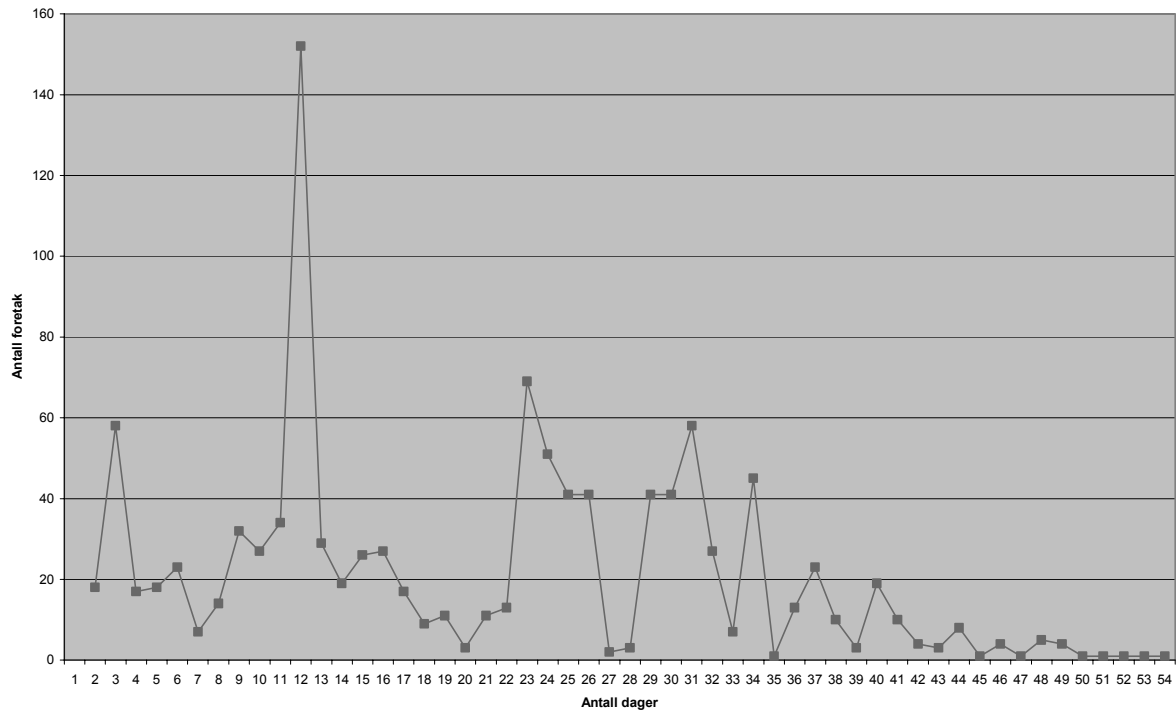


Som man ser av figur 3 gir dette et helt annet bilde av hvilken innrapporteringsform som er mest utbredt. Mer enn 105 000 av de drøyt 160 000 innrapporterte personene i 1. kvartal 2000 ble innrapportert i tråd med SSBs kravspesifikasjon for elektronisk innberetning. I tillegg kommer 30 000 personer som ble innrapportert på excel. Dette vil si at nær 85 prosent av de sysselsatte personene som ble innrapportert til den kvartalsvise lønnsindeksen 1. kvartal 2000 kom på elektronisk form.

Figur 2 og 3 viser også tydelig at små foretak rapporterer på skjema mens store foretak rapporterer på elektronisk form. Det er allikevel viktig å merke seg at for å få laget statistikk basert på et representativt utvalg er det helt essensielt å få med både store, mellom-store og små foretak.

Produksjonsprosessen fra data er kommet i hus til de ligger ferdig på tabell kan måles ved å se på innkviseringsdato i forhold til dato for innlest på tabell. Dessverre eksisterer ikke en slik innlest-dato for foretak rapportert på excel, vi må derfor se bort fra disse i denne sammenhengen. Vi kan imidlertid etter all sannsynlighet anta at produksjonstiden ligger nært opp til produksjonstiden for foretakene som er innrapportert i tråd med SSBs kravspesifikasjon. I 1. kvartal 2000 var det om lag 100 foretak, eller om lag 7 prosent, som rapporterte på excel.

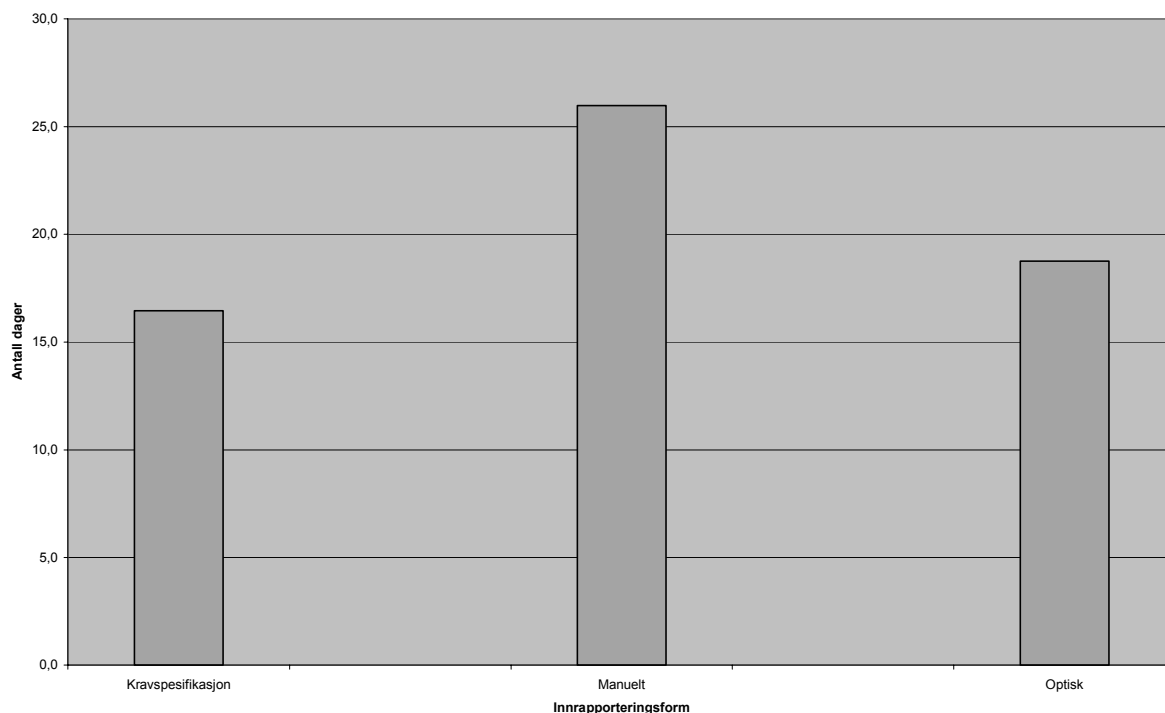
Figur 4: Antall foretak fordelt etter antall dager fra innkwitteringsdato til foretaket er ferdig innlest på tabell



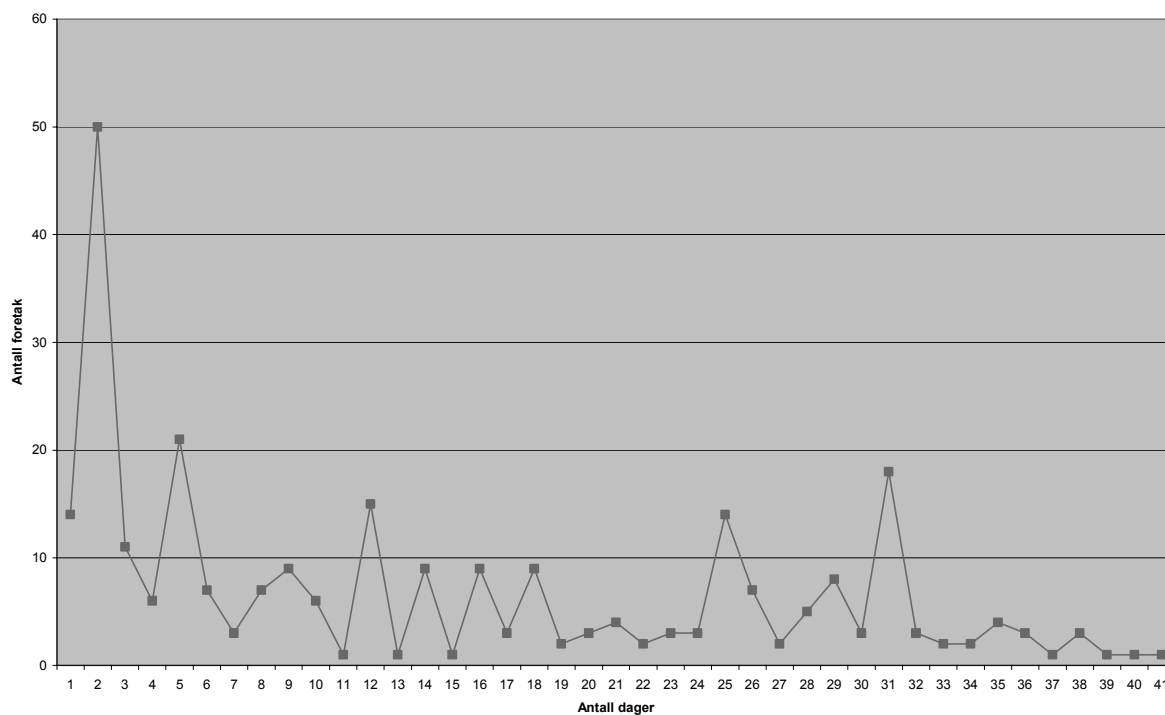
Figur 4 viser at det er stor forskjell i bearbeidingstid fra foretaket er innkvittert til det er ferdig innlest på tabell. En stor andel innkvitteres i løpet av et par uker, men det er også tydelig at det for mange foretak tar atskillig lengre tid. Dette er selve kjernen i problemstillingen, nemlig at bearbeidingen av SSB må ta kortere tid.

Det er nyttig å undersøke hvorvidt innrapporteringsformen har betydning for hvor lang produksjonstid et foretak har.

Figur 5: Antall dager fra et foretak er innkvittert til det er innlest på tabell etter innrapporteringsform



Figur 6: Antall foretak rapportert etter kravspesifikasjon fordelt etter antall dager fra innkviseringsdato til de er innlest på tabell

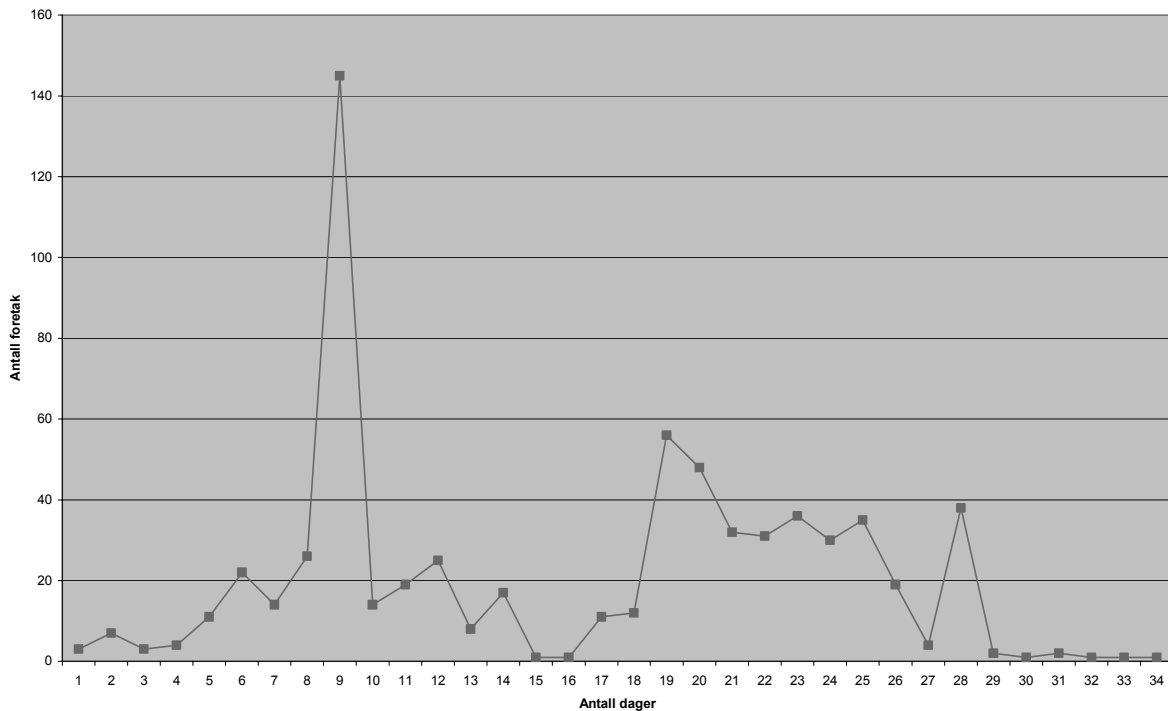


Figur 5 viser at bearbeidingstiden er kortere for foretak som rapporterer elektronisk enn hva tilfellet er for foretak som registreres optisk eller manuelt. At bearbeidingstiden for disketter er lavere enn hva

tilfellet er for foretak som rapporterer på skjema er overraskende, i og med at de store foretakene rapporterer elektronisk og de små foretakene rapporterer på skjema.

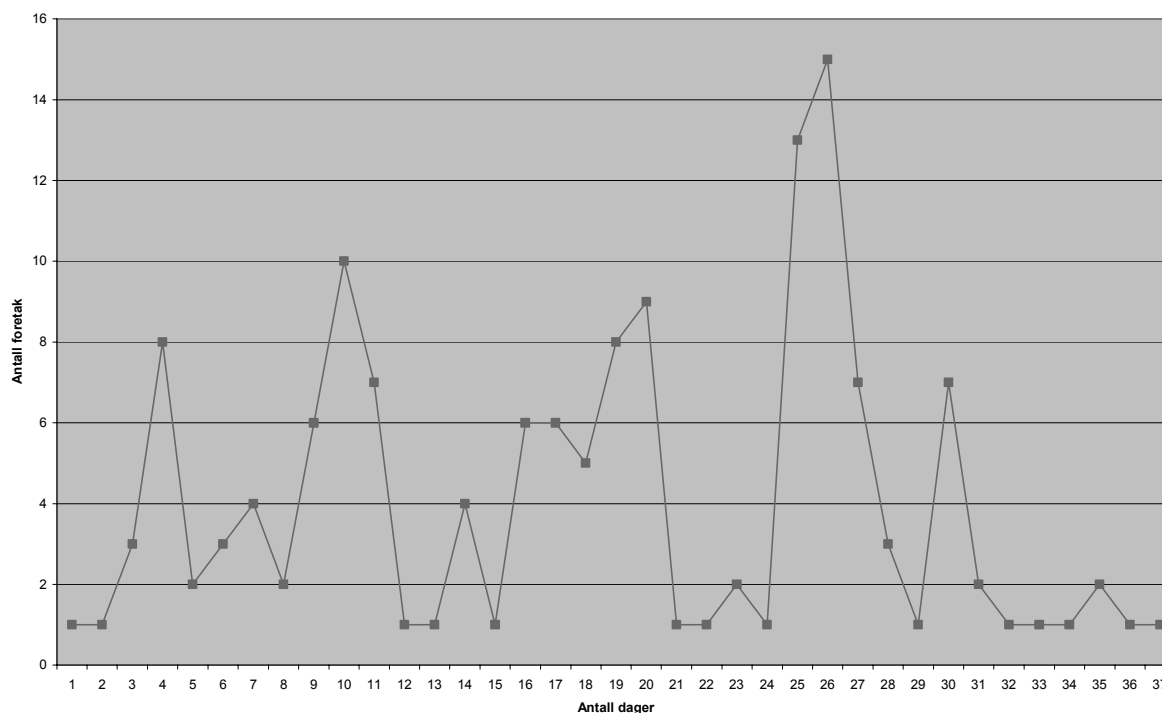
Fra figur 6 er det tydelig at det er et stort sprik når det gjelder bearbeidingstid for foretak som rapporterer i tråd med kravspesifikasjonen. Om lag 75 foretak ble behandlet og innlest på tabell i løpet av et par dager etter at de ble innkvittert, men forøvrig var det veldig forskjellig. At det for mange foretak skal ta over en måned fra innkviseringsdato til de er innlest på tabell er langt fra godt nok.

Figur 7: Antall foretak som er optisk lest fordelt etter antall dager fra innkviseringsdato til de er ferdig innlest på tabell



Også for foretak som rapporterer på skjema er det voldsomt stor sprik i behandlingstiden. Av en eller annen grunn er det knapt noen foretak som blir behandlet før det har gått bortimot en uke. I tillegg er det svært mange som ikke blir innlest på tabell før det har gått 20-30 dager fra de var innkvittert.

Figur 8: Antall foretak manuelt registrert fordelt etter antall dager fra innkwitteringsdato til de er ferdig innlest på tabell



Det er ikke overraskende at foretak som rapporterer på lister og som dermed må registreres manuelt har en lengre bearbeidingstid enn hva tilfellet er for foretak som rapporterer elektronisk eller på skjema som leses optisk. Det virker allikevel som om tidsbruken er i overkant når man ser at det tar 25 dager før tyngdepunktet av denne mengden er innlest på tabell.

Ut fra dette er det tydelig at det er klare forbedringspotensialer for alle innrapporteringsformer. Det aller meste av kontrollene av innkomne data blir under datafangsten foretatt visuelt på skjema/diskett eller ved hjelp av maskinelle kontroller på Oracle-tabell etter at materialet er innlest. Disse kontrollene er imidlertid hovedsakelig fokusert rundt enkeltobservasjoner uten at disse på noen måte blir sammenlignet med tidligere kvartal eller lignende. Isolert sett kan en observasjon se korrekt ut på et gitt tidspunkt ut i fra gitte rammer på sentrale variable, men dette trenger ikke nødvendigvis være tilfelle. Vi må ha en mer omfattende kontroll før vi kan foreta en endelig godkjenning.

Dersom vi på et tidlig stadium skal kunne avsløre feil i et materiale på bakgrunn av sammenligninger med tidligere rapporteringer, bør vi ha en slik kontroll kontinuerlig i prosessen. Et sentralt element i kvalitetssikringen er å avsløre feil og problemer på et så tidlig stadium at vi har mulighet til å få tallene korrigert innen frigivningen. Dersom dette ikke er tilfelle, vil vi bli nødt til å vrake de aktuelle observasjonene, noe som alltid er uheldig.

Måten vi foretar en slik kontroll på er at vi kobler innkomne foretak det aktuelle kvartalet med identiske foretak ved siste årsstatistikk (tilsvarer siste 3. kvartal). Her kan vi i tillegg koble identiske personer og se hvilken lønnsutvikling de har den aktuelle perioden. Som et eksempel kan vi bruke følgende:

En person har i 3. kvartal 20 000 kroner i utbetalt avtalt lønn per måned, mens samme person har 25 000 kroner i utbetalt avtalt lønn per måned i 4. kvartal. Hvert av disse tallene vil isolert sett bli godkjent i alle våre kontroller, men vi vil reagere når vi foretar sammenligning mellom kvartalene. Dette tilsvarer en vekst på 25%, noe som er teoretisk mulig, men så avgjort ikke vanlig. Vi kan da foreta en kontroll av dette ved å kontakte det aktuelle foretaket og da få bekreftet/avkreftet dette.

Dersom foretaket finner ut at det her er rapportert feil, kan vi få tilsendt en ny og korrekt versjon fra foretaket. Dette forutsetter imidlertid at vi oppdager denne feilen på et så tidlig stadium at vi rekker å få en ny versjon som vi kan benytte.

Generelt er det slik at vi ikke kontakter et foretak dersom én person har en slik vekst. Vi tar kontakt med foretaket dersom vi av materialet kan se at det er en trend i foretaket at alle, eller enkelte grupper, har en slik avvikende utvikling.

Essensen i kvalitetsforbedringen er at vi på et så tidlig tidspunkt som mulig kan avsløre feil og mangler med et materiale. Testen vist ovenfor blir i hovedsak utført i den siste produksjonsprosessen, men det er ingen tvil om at den med fordel kan utføres langt tidligere i prosessen.

Følgende punkter kan sies å være sentrale når man skal få en raskere bearbeiding av innkomne data:

- Innkommet materiale må tas tak i med en gang det kommer i hus. Dersom disketter, skjemaer eller lister blir liggende i lengre tid blir prosessen unødvendig lang. Ut i fra gjennomgangen av behandlingstiden for de ulike innrapporteringsformene er det tydelig at dette har vært tilfelle for alle, men mest utbredt blant foretak som kommer inn på skjema eller lister. Vi må regne med at det trengs en viss bearbeidingstid for disketter, som tross alt dekker 85 prosent av totalt antall sysselsatte personer. Til tross for dette er det heller ikke noen tvil om at disketter også kan bli gjenstand for en hurtigere bearbeiding.
- Systemet med at alle disketter blir kontrollert umiddelbart etter å ha kommet i hus for å avdekke feil eller mangler er helt essensielt. Dette medfører at foretaket umiddelbart får tilbakemelding og kan ha mulighet til å foreta korreksjoner på et så tidlig tidspunkt at det fortsatt er mulig å få materialet med i statistikken.
- Det er viktig å ha en kontinuerlig dialog mellom seksjon 420 og seksjon 450. Dersom det er problemer med sentrale foretak er det viktig at seksjon 420 blir informert om dette.
- Seksjon 420 må foreta kontinuerlige kontroller av materialet som kommer inn på tabell. Alle kontroller i innlesingsprosedyrene går isolert på hver enkelt observasjon. Feil kan i enda større grad avdekkes dersom det foretas kontinuerlige kontroller mot tidligere leverte data fra foretaket. Dette innebærer å foreta tabellproduksjonen på langt tidligere tidspunkt enn tidligere for å avdekke feil og mangler. Dersom dette blir gjort kontinuerlig og tidlig i prosessen kan foretaket få tilbakemelding om usammenhengende tall og komme med korrigeringer i tide før frigivning.
- Seksjon 420 må foreta kontinuerlige kontroller av at innkvitterte foretak blir lest inn på tabell. Dersom foretak ikke blir lest inn på tabell selv om de er registrert innkommet er det viktig å avdekke dette på et så tidlig tidspunkt som mulig.
- Det er mange problemer med den optiske lesingen, både tidsmessig og kvalitetsmessig. Prosjektet mener derfor at seksjon 450 bør undersøke hvorvidt kvaliteten på den optiske lesingen til bruk i lønnsstatistikken er god nok. Herunder inngår også en vurdering av om utformingen av skjemaet kan forbedres.

6 Konklusjon

Produksjonsprosessen til den kvartalsvise lønnsindeksen er sammensatt og også relativt komplisert. Mellom 150 000 og 200 000 individopplysninger skal innhentes og bearbeides i løpet av kort tid. Å redusere produksjonstiden med 30 dager krever at data må komme raskere inn til SSB og at SSB får til en hurtigere bearbeiding av dataene.

Det er viktig å være klar over hvilke konsekvenser et krav om en så kraftig tidsbegrensning kan ha:

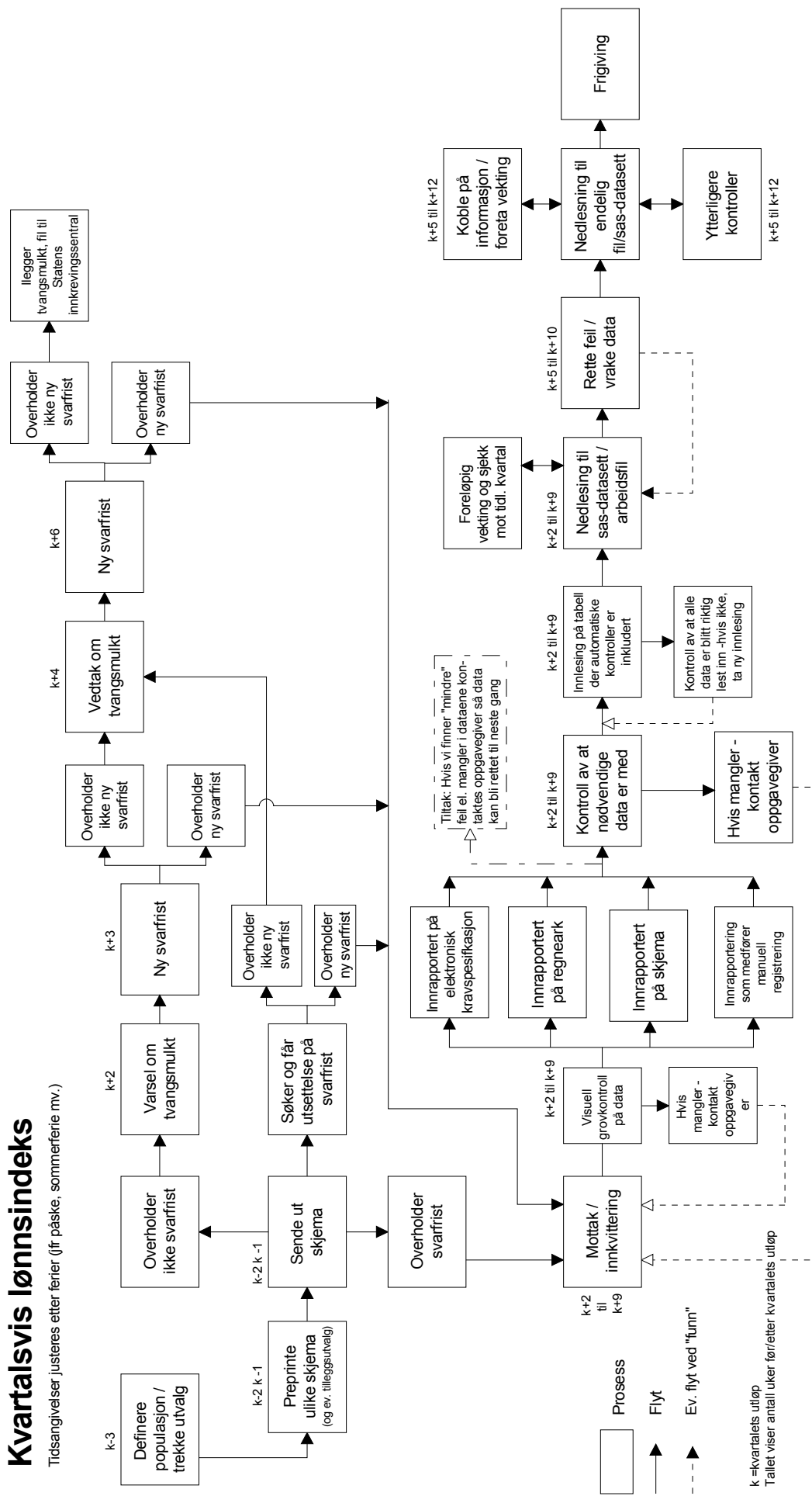
- Statistikken blir mer sårbar. Dersom det er problemer med foretak kan det være for kort tid til å rette det opp, i og med at fristene er så stramme. Fram til nå har det vært litt mer tid til å få inn korrigerte opplysninger fra foretak som har rapportert feil mv. Sannsynligvis blir vi derfor nødt til å godta et noe større frafall enn tidligere som følge av at vi ikke rekker å få inn all informasjon.
- I tillegg er det enda mer alvorlig enn tidligere dersom noe av SSBs tekniske utstyr svikter. Eksempler på dette er scanneren som foretar optisk lesing av skjema. Denne har ved flere anledninger blitt ødelagt, noe som har medført at den ikke har kunnet benyttes før reparatører har fått se på den. Slike problemer blir mer alvorlige når fristene blir kortere.
- Noen næringshovedområder er svært sårbare som følge av at noen få store foretak er helt essensielle for å gi en riktig statistikk. Dersom det blir problemer med noen av disse foretakene er det mindre tid til å få dette korrigert, noe som i ytterste konsekvens kan medføre at enkelte næringshovedområder må holdes tilbake ved frigivningen.

Generelt bør det imidlertid være fullt mulig å gjennomføre statistikken 30 dager raskere enn tidligere, men det krever at rutineene blir fulgt til punkt og prikke. I tillegg er det viktig å påpeke at for at dette skal kunne gjennomføres er det helt sentralt at statistikken gis høy prioritet ved seksjon 420 og at seksjon 450 har nok ressurser til å jobbe med den kvartalsvise lønnsindeksen. Enkelte perioder av året vil det være helt nødvendig at seksjon 450 får muligheten til å jobbe så mye som trengs for å denne statistikken ferdig. Seksjon 450 har kontinuerlig press fra mange seksjoner for å få data ferdig. En så stram innskjerping av produksjonstiden til den kvartalsvise lønnsindeksen medfører at det på enkelte tidspunkt ikke er muligheter til å nedprioritere denne og la datamaterialet bli liggende over lengre tid.

Vedlegg 1: Flytrediagram for produksjonsprosessen

Kvartalsvis lønnsindeks

Tidsangivelser justeres etter ferier (jfr påske, sommerferie mv.)



Vedlegg 2: Formålsbeskrivelse

Seksjon for inntekts- og lønnsstatistikk
Str 01.03.2001

Prosjektnavn: Datafangst lønnsstatistikk

Formålsbeskrivelse

Med bakgrunn i økte krav fra brukerne til aktualitet i kvartalsvis lønnsindeks (fra 90 dager til 60 dager etter tellingstidspunktet), ønsker seksjon 420 en gjennomgang av produksjonsprosessene fra utsending av skjema til ferdige tall på tabell i Oracle. Målsettingen med prosjektet er å belyse alle prosesser i forbindelse med datafangsten og vurdere om noe kan endres for å øke aktualiteten. Eventuelle resultater fra prosjektet vil være relevant for alle lønnsstatistikkene.

Kvartalsvis lønnsindeks omfatter mange prosesser, fra trekking av utvalg, tilrettelegging av skjema, utsending til oppgavegiver, innsamling av data, kontroll av data, innlasting på tabell, produksjon av tabeller og presentasjon. Prosjektet ønsker å fokusere på prosessene fra og med utsending av skjema til oppgavegiver og fram til data er tilrettelagt på tabell og klare til å benyttes i tabellproduksjon. Dette vil si at vi ikke vil gå inn på de første og siste prosessene i arbeidet med kvartalsvis lønnsindeks. Årsaken til dette er at vi ønsker å avgrense prosjektet i størst mulig grad mot det som er målsettingen, nemlig å redusere produksjonstiden i tråd med internasjonale og interne krav.

De sist utgitte publikasjonene i serien Notater

- 2001/38 K.I. Bøe og S. Lien: FD - Trygd: Dokumentasjonsrapport. Attføringspenger. 1992-1999. 80s.
- 2001/39 P.E. Lilleås, og R. Nygaard Johnsen: Foreldrebetaling i kommunale og private barnehager i perioden 1992-2001. 23s.
- 2001/40 B.R. Joneid: KOSTRA GenRev 2000. Malverk for generelt revisjonssystem - KOSTRA-data. Del 1: Håndbok for bruk av malverket ved generering av applikasjoner. Del 2: Systemdokumentasjon av GenRev-malverket. 46s.
- 2001/41 T.M. Normann: Bostedets betydning. Dokumentasjonsrapport. 36.
- 2001/42 B.A. Holth og J.T. Prangerød: Lederskapsundersøkelsen 2000. Dokumentasjonsrapport. 245s.
- 2001/43 T. Fæhn, J.-A. Jørgensen, T. Åvitsland, W. Drzwi: Næringsfordelte skatteutgifter. Dokumentasjon av beregningsgrunnlaget i ERA-beregningene 1998. 47s.
- 2001/44 KOSTRA - VAR-rapport 2001. 34s.
- 2001/45 KOSTRA - Kulturminne, natur og nærmiljø. 38s.
- 2001/46 KOSTRA: Arbeidsgrupperapporter 2001 - hefte 1. 58s.
- 2001/47 KOSTRA: Arbeidsgrupperapporter 2001 - hefte 2. 46s.
- 2001/48 Rapport fra arbeidsgruppa for KOSTRA - Samferdsel. 27s.
- 2001/49 L. Vågane: Undersøkelse om fysisk aktivitet blant personer i alderen 55-75 år: Dokumentasjonsrapport. 22s.
- 2001/50 L-C. Zhang: Estimeringsmetode for familie-/husholdningsfordeling. 17s.
- 2001/51 K.I. Bøe og J. Lajord: FD - Trygd: Dokumentasjonsrapport. Statsansatte. 1992-1999. 28s.
- 2001/52 I. Sagelvmo og H. Sjølie: Beregning av næringene jordbruk og skogbruk i nasjonalregnskapet. 51s.
- 2001/53 L.Solheim: Kvartalsvis lønnsindeks - definisjon av parametere, beregning av estimer og overvåkning av kvaliteten. 24s.
- 2001/54 O. Klungsoyr: Sesongjustering av tids-serier. Spektralanalyse og filt-rering. 47s.
- 2001/55 F. Brunvoll, S. Homstvedt og H. Høie: Mulighetenes marked? SSB-statistikk til regjeringens resultatoppfølging på miljøvernområdet. Potensial og foreløpige prioriteringer. 153s.
- 2001/56 E. Rønning: Trekk ved barn og unges levkår på 1990-tallet. 75s.
- 2001/57 A. Andersen, A.G. Hustoft, A. Rolland, S.T. Vikan: Dokumentasjon av levkårsundersøkelsene. 63s.
- 2001/58 J. Lajord, C. Nordseth: FD - Trygd: Dokumentasjonsrapport. Arbeidssøkere. 1992-1999. 76s.
- 2001/59 A.K. Enge, V. Hansen og B. Tornsjø: Planlegging av et statistikkssystem for energibruk i næringsbygg. 47s.
- 2001/60 G. Daugstad, J. Einarsen, B. Holtet, T. Krokstad og T. Vangen: Dokumentasjonsnotat for FylkesKOSTRA videregående opplæring 2001. 127s.
- 2001/61 J. Epland og M.I. Kirkeberg: Dokumentasjon av inntektsstatistikken for personer og familier 1993-1998: En nærmere beskrivelse av inntektsvariabler for Folke- og boligtellingen 2001. 51s.
- 2001/62 R.N. Johnsen: Undersøking om foreldrebetaling i barnehagar, august 2001. 39s.
- 2001/63 T. Granseth: Formidling av private hytter gjennom hytteformidler. 30s.
- 2001/64 R. Johannessen: Mikroindeksformel i konsumprisindeksen. 24s.