



*Frants Gundersen og Arnt Even Hustad*

# Notater

**Statistikk over anmeldte  
lovbrudd og registrerte ofre**  
Dokumentasjon

# Forord

Statistikken over anmeldte lovbrudd ble første gang utarbeidet i 1990 og har sammenliknbare tall fra og med 1991. Det ble for første gang laget statistikk over registrerte ofre i 1997. Det har tidligere ikke vært utarbeidet og publisert statistikk over registrerte ofre som en del av den løpende statistikken. I løpet av 1999 ble det imidlertid lagt ned et betydelig arbeid for å etablere et datagrunnlag som en slik offerstatistikk kunne baseres på. Det finnes i dag data om registrerte ofre tilbake til 1993.

Dette notatet beskriver den faglige og tekniske bearbeidingen av rådataene, slik disse rutineene har blitt utarbeidet i perioden 1991–1999. Bearbeidingsrutinene er mer et resultat av en kontinuerlig prosess enn et enkeltstående prosjekt, men alle årganger har per i dag gjennomgått de samme bearbeidingsrutinene der ikke annet er spesifisert i notatet.

Nye kjennemerker er inkludert i datamaterialet i perioden. Det er først og fremst flere personopplysninger (landbakgrunn) og en finere inndeling av gjerningssted. Dette gir nye muligheter både i bearbeidingsprosessen og i presentasjonen av statistikken, men forutsetter også et mer omfattende bearbeidingsopplegg.

# Innhold

## Tabell og figuroversikt

Figurer .....	2
Tabeller .....	2
<b>1. Statistikkene</b>	
1.1 Bakgrunn .....	3
1.2 Formål.....	3
1.3 Datagrunnlag .....	4
1.4 Feilkilder .....	5
<b>2. Statistikkproduksjonen</b>	
2.1 Teknisk løsning .....	8
2.2 Om dokumentasjonen.....	8
2.3 Gjennomføring .....	9
<b>3. Bearbeiding</b>	
3.1 Generelt .....	10
3.2 Anmeldte lovbrudd.....	11
3.3 Halvårsstatistikk over anmeldte lovbrudd.....	16
3.4 Registrerte ofre.....	16
3.5 Dokumentasjon av bearbeidingen .....	33
<b>4. Tabellproduksjon</b>	
4.1 Generelt .....	35
4.2 Utvelgelse.....	35
4.3 Begrensninger i bruk av variable.....	36
<b>5. Permanent lagring av data</b>	
5.1 Lagring av ferdig revidert materiale.....	37
5.2 Lagring av andre filer permanent .....	40
5.3 Innlesing av ascii-filer til SAS-datasett.....	40
<b>6. Referanser</b> .....	42
<b>7. Vedlegg</b>	
7.1 Filstruktur .....	43
7.2 Kataloger .....	45
7.3 Kontroller i STRASAK.....	48
7.4 Spesielle variabelverdier .....	49
7.5 Bruk av fellesressurser på UNIX.....	50
De sist utgitte publikasjonene i serien Notater.....	51

# Tabell og figuroversikt

## Figurer

Figur 1.	Et lovbrudds gang gjennom rettsapparatet .....	4
Figur 2.	Andelen personer og registrerte ofre med gyldig fødselsnummer før og etter bearbeiding. 1993–1998 .....	7
Figur 3.	Logisk dataflyt, anmeldte lovbrudd.....	12
Figur 4.	Logisk dataflyt, registrerte ofre, del 1 .....	17
Figur 5.	Logisk dataflyt, registrerte ofre, del 2 .....	18
Figur 6.	Logisk dataflyt, registrerte ofre, del 3 .....	19
Figur 7.	Beregning av alder.....	30
Figur 8.	Logisk dataflyt, arkivering og innlesing av årgangsdata .....	37

## Tabeller

Tabell 1.	Filbeskrivelse av de enhetene fra rådataene som viser saksopplysninger .....	13
Tabell 2.	Beskrivelse av det ferdig bearbeidede materialet, anmeldte lovbrudd.....	16
Tabell 3.	Filbeskrivelse av enhetene fra rådataene som viser saksopplysninger.....	20
Tabell 4.	Filbeskrivelse av enhetene fra rådataene som viser personopplysninger .....	20
Tabell 5.	Koblingene mot BEBAS, kjennemerker og risiko for feilkobling .....	26
Tabell 6.	Filbeskrivelse av BBAS[ <b>aar</b> ].ssd04 og BBAS[ <b>aar</b> +1].ssd04 .....	27
Tabell 7.	Filbeskrivelse av \$KRIMINAL/fodeland/g[ <b>aar</b> ].txt.....	28
Tabell 8.	Filbeskrivelse av det ferdig bearbeidede offermaterialet, g[ <b>aar</b> ].ssd04 .....	33
Tabell 9.	Utvelgelse ved bruk av forskjellige enheter i tabellene.....	35
Tabell 10.	Variable med begrenset bruksområde .....	36
Tabell 11.	Filbeskrivelse av årgangsfilene ved lagring av flat fil – anmeldte lovbrudd.....	38
Tabell 12.	Filbeskrivelse av årgangsfilene ved lagring av flat fil – registrerte ofre.....	39

# 1. Statistikkene

## 1.1 Bakgrunn

Statistikk over anmeldte lovbrudd ble publisert første gang i NOS Kriminalstatistikk 1990. Den gang var ikke alle politikamrene tilkoblet STRASAK, slik at årgangen 1991 regnes som den første komplette. Sysselmannen på Svalbard ble imidlertid ikke tilknyttet før 1996. Til å begynne med foregikk all bearbeiding på stormaskin. Senere er all bearbeiding utarbeidet i SAS under UNIX. Denne dokumentasjonen inneholder derfor ingen informasjon om bearbeidingen slik den var på stormaskin.

Statistikk over anmeldte lovbrudd og registrerte ofre er slått sammen i et felles dokumentasjonsnotat. Dette fordi de to produseres ut fra samme råmateriale. Statistikkene har derfor også en del felles kjennemerker og bearbeidingsrutiner. Statistikkene er basert på politiets "råmateriale", det vil si første stadium når et lovbrudd registreres i et administrativt register. Dette gir muligheter for rask rapportering og høy aktualitet på statistikken. Men det betyr også at kvaliteten på materialet ikke er etterprøvd, siden politiet enda ikke har foretatt etterforskning på basis av anmeldelsene. Det innkomne datamaterialet vil derfor være av noe mer usikker kvalitet enn på andre deler av kriminalstatistikken.

Dette notatet beskriver den faglige og tekniske bearbeidingen av rådataene. Denne bearbeidingen ble utarbeidet i perioden 1995–1999. I tillegg dokumenteres følgende punkter som er kommet til parallelt eller i etterkant av utarbeidingen av bearbeidingsrutinene:

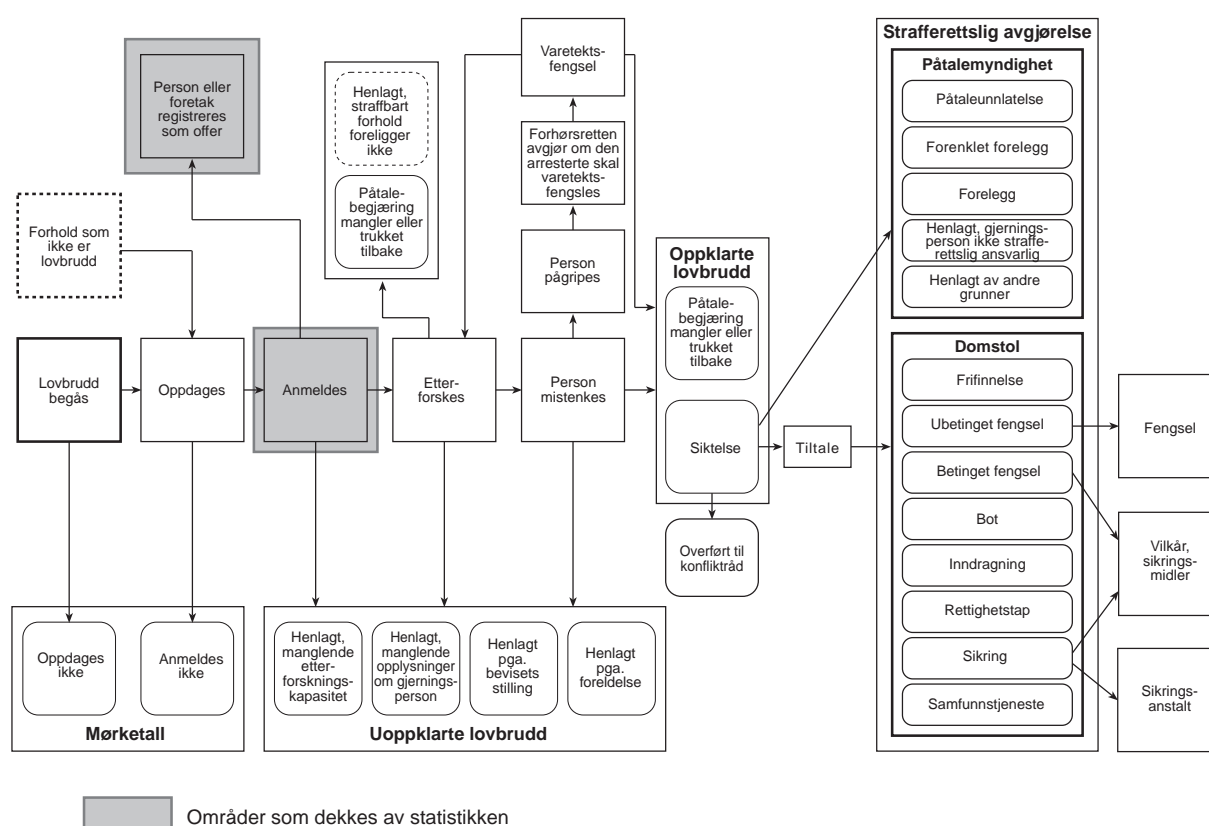
- Eldre årganger er konvertert til samme format for å kunne lage tidsserier. Første årgang er 1990 for anmeldte lovbrudd, mens 1993 er første årgang for registrerte ofre. Ikke alle kjennemerker er gyldige for alle år, men kapittel 4.3 gir en oversikt over brudd i tidsseriene.
- Nye kjennemerker er inkludert i datamaterialet (sonen der lovbruddet ble begått og landbakgrunn for registrerte ofre) som gir nye muligheter i bearbeidingsprosessen.

## 1.2 Formål

Statistikken over anmeldte lovbrudd blir publisert fire til seks uker etter statistikkperiodens utløp, og statistikkområdet med høyest aktualitet innen kriminalstatistikken. I form av relativt rask rapportering er Statistisk sentralbyrå i stand til å kunne si noe om det registrerte kriminalitetsnivået fra år til annet, både på nasjonalt nivå og etter fylke, politidistrikt, kommune. Statistikk kan også lages på bydelsnivå for Oslo og Bergen ved hjelp av opplysning politikammer og sone. Statistikken kan også benyttes som indikator for kriminalitetsutviklingen for ulike typer lovbrudd, fra 1990/1991 og fram til i dag. Datamaterialet representerer klager fra publikum som er upåvirket av andre forhold, som for eksempel etterforskningskapasiteten ved det enkelte politidistrikt. For noen typer lovbrudd, slik som narkotika- og trafikkriminalitet, er sakstilfanget sterkt avhengig av politiets egne prioriteringer.

Sett i forhold til brukerens behov, er det ulike kvalitetskrav for forskjellige deler av kriminalstatistikken. For anmeldte lovbrudd er aktualitet aller viktigst. Derfor produseres denne statistikken både for første halvår og for hvert år. Fram til 1998 ble det også produsert kvartalsvis statistikk. Denne har blitt lagt ned, hovedsakelig på grunn av tilfeldige variasjoner forårsaket av svært korte statistikkperioder. Dette understreker betydningen av at statistikken bør leses som trender over lengre tid for å kunne være substansiell i forhold til kriminalitetsutviklingen på gitte områder.

Offerstatistikken har kvaliteter i seg som indikator på ulike gruppers utsatthet for kriminalitet. Siden mørketallene er store er statistikken først og fremst egnet til å sammenligne grupper, og å følge grupper over tid. Statistikken vil i mindre grad kunne vise den faktiske utsattheten for kriminalitet ved et gitt tidspunkt. Figur 1 viser hvilke deler av rettsapparatet som statistikkene beskriver.



Figur 1. Et lovbrudds gang gjennom rettsapparatet

### 1.3 Datagrunnlag

Datagrunnlaget er fra politiets sentrale registreringssystem, STRASAK, som dekker alle registrerte lovbrudd ved landets politidistrikter. Det er Politiets datatjeneste (PD) som står for utkjøringen av datamaterialet som leveres til Statistisk sentralbyrå på tape. Driftsavdelingen ved Statistisk sentralbyrå legger materialet over til spesifisert sted på UNIX.

Hvert registrerte lovbrudd utgjør vanligvis én enhet i materialet. Lovbruddet er kjennetegnet med unik verdi for politidistrikt, registreringsår og anmeldelsesnummer. For en videre beskrivelse av

datagrunnlag, kjennemerker og definisjoner henvises det til Statistisk sentralbyrå (1998b), del 1 og del 2.

STRASAK inneholder noen gyldighetskontroller av data. En oversikt over disse finnes i vedlegg 7.3. Kontrollene sikrer kvaliteten på rådata, og kvaliteten er generelt bedret siden første produksjonsår. Etterhvert som det blir innarbeidet nye kontroller i STRASAK kan ytterligere forbedring av det innkomne materialet forventes.

Med hjemmel i statistikklovens §3-2 (1) har Statistisk sentralbyrå rett til å benytte STRASAK som grunnlag for offisiell statistikk. I henhold til statistikkloven §3-2 (2) skal alle større endringer av STRASAK meldes til Statistisk sentralbyrå før iverksettelse. Myndighet for håndhevelse av statistikkloven §3-2 er delegert til Finans og tolldepartementet, mens Statistisk sentralbyrå selv har myndighet til å gi pålegg om oppgaveplikt etter statistikklovens §2-2.

## 1.4 Feilkilder

Datamaterialet fra Politiets datatjeneste er en fulltelling av registrerte enheter, slik at frafall på enhetsnivå utelukkes. Statistisk sentralbyrå har imidlertid ingen mulighet til å kontrollere frafall av enheter ved manglende registrering i de enkelte politidistrikt. I tillegg til manglende registrering av kjente lovbrudd vil det være mange lovbrudd som politiet aldri får kjennskap til. Mørketallene vil variere med lovbruddstype, der de minst alvorlige lovbruddene generelt har størst mørketall<sup>1</sup>. Betegnelsene *registrerte/anmeldte lovbrudd* og *registrerte ofre* benyttes ofte for å presisere hvilket datamateriale som ligger til grunn for statistikkene.

Partiell frafall (ved at enkelte kjennemerker ikke er oppgitt på enkelte enheter) forekommer, men ved hjelp av interne kontroller i STRASAK er disse begrenset til mindre sentrale kjennemerker, slik som bokommune og statsborgerskap. Kontrollene i STRASAK er listet opp i vedlegg 7.3.

For at STRASAK skal fungere i de daglige rutinene ved de enkelte politidistriktene er ikke alle kontrollene som er innebygd absolutte, men av mer veiledende karakter. Rutinene for bearbeidingen (kapittel 3) kan derfor ikke baseres på at disse kontrollene fungerer absolutt.

For politiet kan det være vanskelig å avgjøre om en straffbar handling skal registreres som ett eller flere lovbrudd. Derfor ble det sommeren 1994 utarbeidet et direktiv for politiets registrering av lovbrudd (Justisdepartementet 1994). Dette sier at når anmeldelsen omfatter flere lovbrudd, skal alle lovbrudd som ut fra en forstandig vurdering kan komme til å utgjøre et eget punkt i siktelsen, registreres. De fleste politidistrikt endret registreringspraksis allerede i 1995, andre senere. Før denne innføringen var praksis at det bare ble registrert ett lovbrudd når

- Flere lovbrudd var begått ved en handling mot ett offer. Det groveste ble registrert.
- Flere ensartede lovbrudd var begått av samme person i et tidsrom mot ett offer.
- Ett lovbrudd var begått ovenfor en gruppe ofre
- Det ene lovbruddet var begått for å bane vei for det andre.

---

<sup>1</sup> For en del lovbrudd kan andre faktorer være helt avgjørende for hvorvidt lovbruddet blir anmeldt eller ikke, slik som dokumentasjon overfor forsikringsselskap, søknad om voldsskadeerstatning, lovbruddets synlighet, relasjon til gjerningsperson og så videre. Mørketallsproblematikk spiller imidlertid først og fremst inn ved tolkning av statistikken, og ikke ved den tekniske vurderingen av datamaterialet.

Det ble registrert flere lovbrudd når et lovbrudd ble begått for å skjule et annet.

Endringene ga et hopp i registrerte lovbrudd fra 1994 til 1995. Økningen skyldtes ikke alene en økning i den faktiske kriminaliteten. Det er rimelig å anta at flere lovbrudd, som tidligere ikke ville ha kommet med i statistikken fordi de skjedde i saker der mer alvorlige lovbrudd også ble begått, nå ble registrert av politiet. Hvor mange dette dreier seg om er det ingen mulighet til å undersøke, siden de nye retningslinjene ble innarbeidet i registreringspraksisen ved ulike tidspunkt ved de ulike politikamrene og tidligere registreringspraksis ikke er dokumentert. Problematikken er blant annet omtalt i Olaussen (1996).

Statistikkene er basert på politiets "råmateriale", det vil si første stadium når et lovbrudd registreres. Det betyr blant annet at politiet ikke har etterprøvd kvaliteten på materialet gjennom eget arbeid. Politiets arbeid kan gi endringer i en liten andel av lovbruddene fordi:

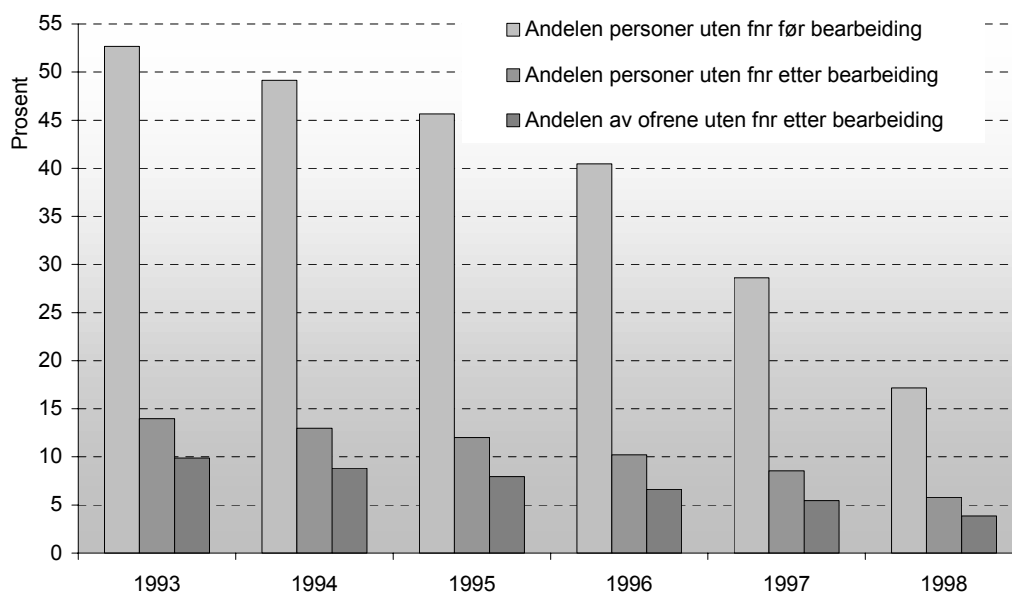
- noen saker (0–3 prosent, avhengig av lovbruddstype) blir henlagt fordi politiet konkluderer med at lovbrudd ikke fant sted eller at det ikke kan bevises at et lovbrudd fant sted. Saken kommer derfor ikke med i statistikken som dekker et senere stadium av straffesakskjeden.
- lovbruddskoden kan bli endret etter hvert som politiet får bedre kjennskap til saken.
- variable knyttet til personer involvert endres og kompletteres

Det innkomne datamaterialet vil derfor være av noe mer usikker kvalitet enn andre deler av kriminalstatistikken.

Materialet er i utgangspunktet ikke komplett med hensyn til personopplysninger til offeret. I råmaterialet for 1998 mangler ca. 17 prosent gyldig fødselsnummer, men gjennom koblinger til Det sentrale folkeregisteret (DSF), basert på andre opplysninger (navn, fødselsdato og adresse), påføres en del av disse gyldig fødselsnummer.

Det er ca. 4 prosent av personene, det vil si cirka 8 000, som ikke kan påføres gyldig fødselsnummer gjennom slik kobling. Av disse kan litt over 2 000 identifiseres til å bo i utlandet (turister som utsettes for lovbrudd i Norge) og 1 500 har uoppgitt bosted (vet ikke om de tilhører befolkningen i Norge eller er utenlandske turister). Vi har da en rest på litt under 5 000 som vi ikke får identifisert i materialet, men som har oppgitt bosted i Norge. Disse fordeler seg jevnt med hensyn til kjønn og alder, men er noe oftere å finne i kategorien *økonomisk kriminalitet* og med gjerningssted/bosted Oslo. Datamaterialet for 1998 har foreløpig den høyeste andelen oppgitte fødselsnumre av de årgangene som Statistisk sentralbyrå har produsert. Figur 2 viser hvordan utviklingen av datakvaliteten var i perioden 1993–1998.





**Figur 2. Andelen personer og registrerte ofre med gyldig fødselsnummer før og etter bearbeiding. 1993–1998**

Ifølge politiets instruksjoner skal verge/foresatte til offeret også registreres, hvis offeret er under myndighetsalder. Verge/foresatte skal også stå som offer. I rutine 06 i kapittel 3.4 skilles disse vergene/foresatte ut. Imidlertid finner rutinen færre verges/foresatte enn antallet umyndige ofre. Dette tyder på en svakhet ved registreringspraksisen ved de enkelte politikamrene, uten at dette nødvendigvis gir dårligere kvalitet på statistikken over de reelle ofrene.

## 2. Statistikkproduksjonen

### 2.1 Teknisk løsning

SAS er valgt som sentral, teknisk løsning i Statistisk sentralbyrå, både for bearbeidingen og lagring av data. Dette er en følge av Statistisk sentralbyrås overgang fra stormaskin til UNIX-baserte tjenere i nettverk. En databaseløsning fra Oracle for lagring er vurdert, men disse statistikkene tjener i liten grad på å bli bygd opp som relasjonsdatabaser.

Statistikken inneholder svært sensitive data og behandles i henhold til *Sikkerhetshåndboka* (Statistisk sentralbyrå 1998a), spesielt kapittel 4 og 9. Krypteringsrutiner for persondata i SAS-format på UNIX er utviklet på sentralt hold i Statistisk sentralbyrå, men er per dags dato ikke implementert i disse statistikkrutinene. Sikring mot misbruk av data ivaretas av Statistisk sentralbyrås generelle restriksjoner på brukertilgang på de stedene i nettverket som dataene fysisk befinner seg. Dette gjøres på operativsystemnivå. Det ligger ingen begrensninger i anvendelsen av dataene hvis brukeren har tilgang til området for kriminalstatistikk (som er '\$KRIMINAL' på UNIX-tjeneren).

Alle årganger av statistikken, både rådata fra Politiets datatjeneste og det ferdig reviderte datasettet, blir lagret permanent som ren ascii-tekst. Dette for å forhindre at eventuelle framtidige skifte av verktøy for databehandling vil medføre tap av data. Dette er nærmere behandlet i kapittel 5.

### 2.2 Om dokumentasjonen

Bearbeidingen og dokumentasjonen av statistikkene følger delvis de retningslinjene som gis blant annet av Berby (1999) og Statistisk sentralbyrå (1995, 1997 og 1998c). Produksjonsrutinene for statistikkene er imidlertid ikke kommet som følge av et avgrenset prosjekt, slik retningslinjene nevnt over tar utgangspunkt i, og dokumentasjonen følger dermed ikke 100 prosent disse retningslinjene. Dokumentasjonen må ses på som en oppsummering av et kontinuerlig arbeid som har foregått over flere år, der det har vært flere personer som har hatt ansvaret for både spesifikasjonene og implementeringen.

All dokumentasjon finnes i elektronisk form (se vedlegg 7.1), og denne vil bli oppdatert av de personene som til en hver tid har ansvaret for statistikkrutinen. Dokumentasjonen skal dekke alle tilgjengelige årganger. Eventuelle endringer ved senere årganger må ikke medføre endring av dokumentasjon som berører tidligere årganger av statistikken. Notatet er også tilpasset tidligere dokumentasjon på kriminalstatistikk ved Statistisk sentralbyrå (Gundersen 1999). Dette gjør at brukere lettere kan ta i bruk flere statistikkområder, og bidrar også til at videre dokumentasjonsarbeid forenkles.

En del av informasjonen (spesielt vedleggene) i dette notatet vil bli lagt inn i Statistisk sentralbyrås DataDok-system. Informasjonen vil da være tilgjengelig elektronisk, og oppdateringer vil kunne gjøres fortløpende (se Bjørsvik og Torvbråten 1999). Statistikken over anmeldte lovbrudd har imidlertid ikke på nåværende tidspunkt (desember 1999) tatt i bruk dette systemet. Hvorvidt

statistikken over registrerte ofre skal ta i bruk systemet vil bestemmes av om statistikken i det hele tatt skal videreføres.

Dokumentasjonen er primært rettet mot to brukergrupper. Den ene er de som i framtiden skal lage og videreutvikle statistikken, den andre brukergruppen er personer som ikke selv er med på bearbeidingen av rådataene, men tar i bruk det ferdige datasettet. Dokumentasjonen er av teknisk karakter, men en del av de faglige vurderingene som ligger bak rutinene i kapittel 3 er også tatt med. Dokumentasjonen er dermed ikke mer teknisk enn at også personer som ikke er direkte involvert i selve databearbeidingen likevel kan ha utbytte av å lese notatet.

## 2.3 Gjennomføring

STRASAK-registeret er fullstendig oppdatert sentralt i løpet av 4–5 dager etter registreringen lokalt, slik Statistisk sentralbyrå skal kunne innhente data 10 til 14 dager etter statistikkperiodens utløp.

Seksjon 320 skal levere "fødelandsfil" og befolkningsdata for kobling i rutine 03 og 04 i kapittel 3.3. Erfaringsmessig tar det tid å skaffe til veie disse dataene, og kontakt med seksjonen bør opprettes i god tid med tanke på å få lesetillatelse til aktuelle filer, kopiere aktuelle filer til kriminalstatistikkenes diskområde eller på annet vis skaffe til veie de nødvendige datasettene (ofte benyttes tilsvarende datasett av flere statistikker).

For statistikk over registrerte ofre er det ikke innarbeidet faste rutiner for publisering.

Gjennomføringen av bearbeidingen av datamaterialet vil i praksis være å følge punktene i bearbeidingsplanen i den rekkefølgen som beskrevet i kapittel 3.

I andre uken av januar kan data fra Politiets datatjeneste (PD) bestilles. I praksis kjenner både fagpersonalet og driftspersonalet ved PD til rutinen, men fagansvarlig ringer vanligvis til våre kontaktpersoner som per 15.12.99 er Dahle eller Kjørstad.

Adressen til PD er:  
Sørkedalsveien 27b  
Postboks 8031 Dep.  
0030 Oslo  
tlf.: 22 06 80 00  
faks.: 22 06 80 80

Gjennomføringen av bearbeidingen er visualisert i figur 2. Symbolbruken for logisk dataflyt finnes i Statistisk sentralbyrå (1997).

Årgangfil for anmeldte lovbrudd kan produseres på én dag. For registrerte ofre er det en del tunge SAS-programmer som i praksis må kjøres om natten. Minst én uke på man regne med fra rådataene er kommet fra PD til ferdig årgangfil kan være ferdig. Imidlertid vil aktualiteten til offerstatistikken først og fremst begrenses av når befolkningsdataene for utløpet av statistikkåret kan være ferdige. Erfaringsmessig vil befolkningsdataene være tilgjengelig i mars–april året etter statistikkåret.

## 3. Bearbeiding

### 3.1 Generelt

Bearbeidingsrutinene er laget slik at det lar seg gjøre å kjøre alle SAS-programmene etter hverandre, fra innlesing av rådata til oppretting av ferdig årgangsfil, uten å behøve å gjøre manuell revisjon eller ekstern kopling, forutsatt at datagrunnlaget for de eksterne koplingene er etablert. Dette betyr at man kan ta hensyn til endringer i bearbeidingen på et hvilket som helst stadium i bearbeidingsprosessen, helt fram til like før årgangsfilen skal være endelig. For å få til dette er blant annet følgende gjort:

- Manuell revisjon er utelatt. Valg og vurderinger er i sin helhet lagt inn i SAS-programmene. Dette er gjort både for å spare tid, men også fordi dataene er "rådata" og dermed så usikre på mikronivå at en manuell revisjonsrutine ville blitt for ressurskrevende i forhold til en eventuell kvalitetsheving som ville finne sted.
- Alle kategoriseringer/omkodinger foregår ved hjelp av SAS-formater, *der SAS-formatene opprettes på nytt for hver programkjøring*. Definisjon av SAS-formatene ligger i selve programmet, mens det er de underliggende katalogene som eventuelt oppdateres.

For alle SAS-programmene opprettes det en logg over bearbeidingen. Denne loggen inneholder informasjon om antall omkodinger, slettinger, innlesinger, treff/ikke treff o.l. som skjer i de enkelte rutinene. Etter at en årgang av statistikken er ferdig bearbeidet vil loggene til sammen danne en omfattende dokumentasjon over den bearbeidingen som er gjennomført.

Dokumentasjonen av den maskinelle rettingen gir muligheter for å vurdere kvaliteten på dataene. Fulgt over tid gir den også mulighet for å avdekke endringer i kvalitet, endringer i registreringsrutiner og eventuelle endringer i STRASAK-registeret som ikke er innarbeidet i bearbeidingen.

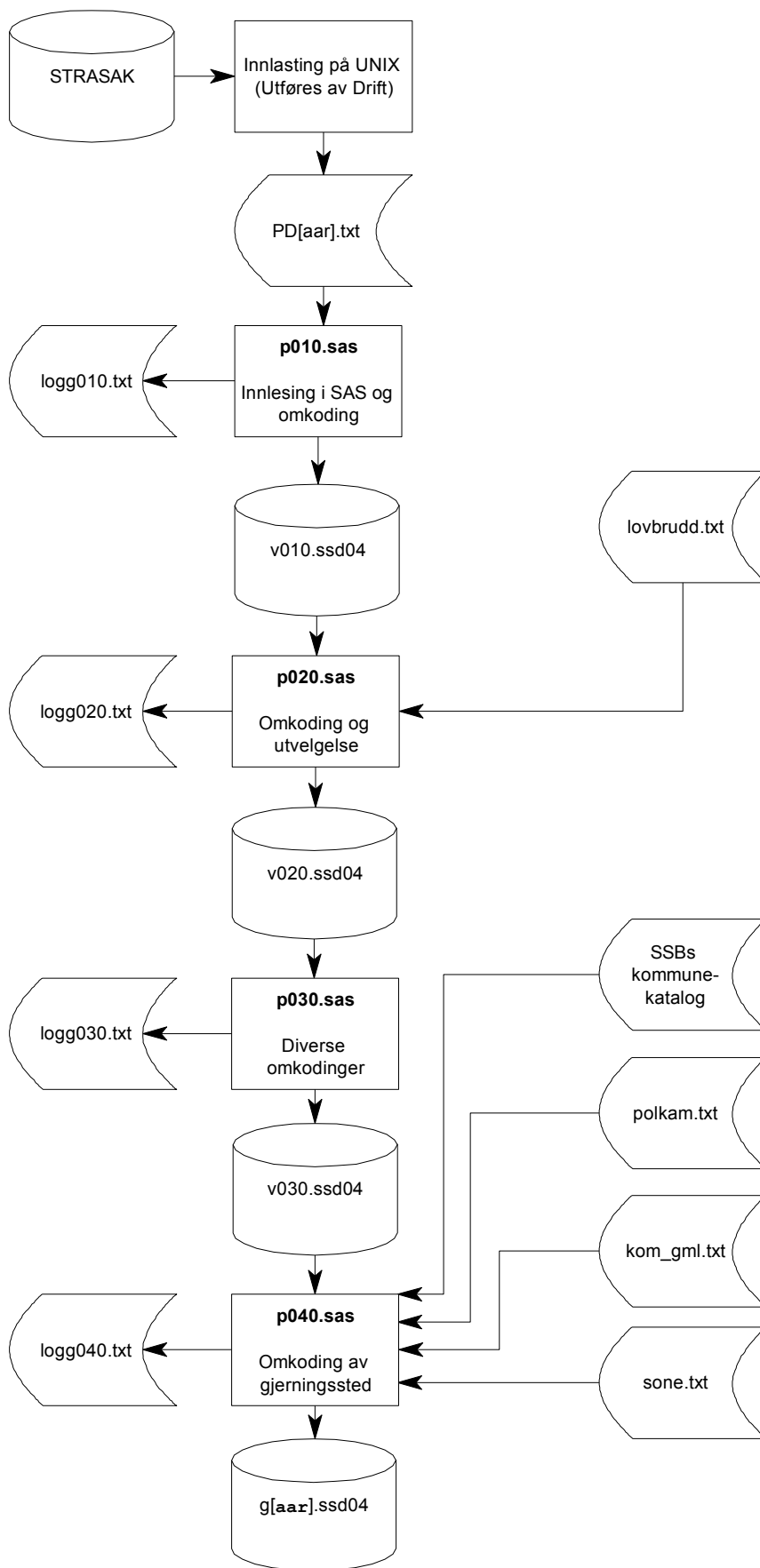
Revisjonen av data følger *Håndbok i datarevisjon* (Statistisk sentralbyrå 1998d) i stor grad, både med hensyn til type kontroller og gjennomføring. Noe revisjon har et element av "automatisk imputering" (se Løvkvist 1997), dvs. at variabelen tilordnes en verdi etter resten av materialet. Den automatiske tilordningen er imidlertid kun basert på verdier til variable på samme enhet, og ikke basert på statistiske fordelinger på hele eller deler av materialet.

I beskrivelsen av filplassering er ofte notasjonen `[aar]` brukt. `[aar]` refererer til variabelen som angir statistikkåret, se for eksempel rutine 01 i kapittel 3.2. Statistikkåret skrives med fire siffer, noe som betyr at datasettet/dokumentet skal ha navn etter hvilken årgang av statistikken som er under bearbeiding. For eksempel vil 'g[aar].ssd04' bety at datasettet skal ha navnet 'g1999.ssd04' hvis det er 1999-årgangen som bearbeides. Tilsvarende vil 'g[aar].txt' gi 'g1999.txt' som filnavn.

## 3.2 Anmeldte lovbrudd

Bearbeidingen av data over anmeldte lovbrudd og registrerte ofre er holdt adskilt. Dette gjøres fordi det bare er råmaterialet som er felles for statistikkene. Det er samme type filstruktur på alle områdene av kriminalstatistikken, og det er SAS-programmer med likelydende navn på områdene (se figurene 3–6). For entydig identifisering av programmene må derfor også filstrukturen eller beskrivelsen av programmene i dette kapitlet benyttes.

Å lage statistikk for anmeldte lovbrudd er i praksis å følge dataflyten i figur 3. Navn på datasett, SAS-programmer og andre dokumenter korresponderer med de beskrevne rutinene i det som følger av dette kapitlet.



Figur 3. Logisk dataflyt, anmeldte lovbrudd

## 01 Innlesning i SAS-datasett

SAS-program: \$KRIMINAL/anmeldt/prog/bearbeid/p010.sas

Inndatasett: \$KRIMINAL/anmeldt/inndata/PD[**aar**].txt

Utdatasett: \$KRIMINAL/anmeldt/arbeid/v010.ssd04

Loggfil: \$KRIMINAL/anmeldt/logger/logg01.txt

A. Når materialet kommer inn, leses det inn i et SAS-datasett fra posisjoner til variable, slik det fremgår av tabell 1. For endelig filbeskrivelse av anmeldte lovbrudd, se tabell 2.

B. Bare enheter der **type** = 'SA' beholdes, det vil si at vi siler ut SA-sakene fra saker merket 'KL' og 'AN'.

C. Det opprettes datovariabler med datoformat i SAS.

**Tabell 1. Filbeskrivelse av de enhetene fra rådataene som viser saksopplysninger**

Posisjon	Beskrivelse	Kortnavn	Format
1–2	Enhetsstype ('SA' = saksopplysninger, 'AN' og 'KL' = Personopplysninger)	<b>Type</b>	Char2
3–4	Politikammer	<b>Polkam</b>	Char2
5–6	Registreringsår	<b>Aar</b>	Num2
7–12	Saksnummer	<b>Saksnr</b>	Char6
13–14	Saksbokstav	<b>Delsak</b>	Char2
15–18	Sekvens	<b>Sekvens</b>	Char4
19–26	Registreringsdato	<b>Regdato</b>	Num8
19–22	Registreringsår	<b>Regaar</b>	Num4
23–24	Registreringsmnd	<b>Regmnd</b>	Num2
25–26	Registreringsdag	<b>Regdag</b>	Num2
27	Statistikkbokstav	<b>Statbok</b>	Char1
28–31	Lovbrudd	<b>Lovbrudd</b>	Char4
32–35	Modus	<b>Modus</b>	Char4
36–39	Sone	<b>Sone</b>	Char4
40–43	Gjerningskommune	<b>Gjkom</b>	Char4
44–51	Gjerningsdato start på handling	<b>Gjdato1</b>	Num8
44–47	Gjerningsår start	<b>Gjaar1</b>	Num4
48–49	Gjerningsmåned start	<b>Gjmnd1</b>	Num2
50–51	Gjerningsdag start	<b>Gjdag1</b>	Num2
52–59	Gjerningsdato stopp	<b>Gjdato2</b>	Num8
52–55	Gjerningsår stopp	<b>Gjaar2</b>	Num2
56–57	Gjerningsmåned stopp	<b>Gjmnd2</b>	Num2
58–59	Gjerningsdag stopp	<b>Gjdag2</b>	Num2

## 02 Omkodning og utvelgelse

I Kriminalstatistikken opererer man ofte med benevnelsen lovbruddsgruppe, basert på inndeling i ti ulike lovbruddsgrupper (tideling). Økonomisk kriminalitet er en slik gruppe, voldskriminalitet en annen. Program 020 oppretter variabel for tideling og blanker ut ugyldige datoer. Videre settes manglende gjerningskommune som blank og **statbok** skal være enten 'K' (forbrytelse) eller 'F' (forseelse). Det betyr at observasjoner merket 'U' (undersøkelsessaker) blir utelatt i den videre bearbeidingen. Lovbruddskoden skal være under '9200'. Observasjoner med høyere kode representerer undersøkelsessaker, det vil si saker som ikke skal telles som anmeldte lovbrudd i statistikken. Saksbokstav skal være blank. Det betyr at vi i det videre utelater delsakene fra datasettet. Politidistrikt skal ikke være lik '56', siden dette er Kripos.

SAS-program: \$KRIMINAL/anmeldt/prog/bearbeid/p020.sas  
Inndatasett: \$KRIMINAL/anmeldt/arbeid/v010.ssd04  
Utdatasett: \$KRIMINAL/anmeldt/arbeid/v020.ssd04  
Loggfil: \$KRIMINAL/anmeldt/logger/logg02.txt

- A. **Tidel** tilordnes verdi etter lovbruddskatalogen (vedlegg 7.2) på grunnlag av **lovbrudd**.
- B. Hvis **gjmnnd1** < 1 eller **gjmnnd1** > 12 settes **gjmnnd1** lik blank. Hvis **gjaar1** < 70 settes **gjaar1** lik blank. Hvis **gjaar1** = . settes **gjmnnd1** lik blank. Hvis **gjmnnd2** < 1 eller **gmnd2** > 12 settes **gjmnnd2** lik blank. Hvis **gjaar2** < 70 settes **gjaar2** lik blank. Hvis **gjaar2** = blank settes **gjmnnd2** lik blank.
- C. Hvis **gjkom** = '0000' settes **gjkom** lik blank.
- D. Hvis **statbok** er merket 'K' eller 'F' og **saksbok** lik blank og **lovbrudd** < '9200' og **polkam** ikke lik '56', beholdes observasjonene.

## 03 Omkodning av statistikkbokstav

SAS-program: \$KRIMINAL/anmeldt/prog/bearbeid/p030.sas  
Inndatasett: \$KRIMINAL/anmeldt/arbeid/v020.ssd04  
Innfil: \$KRIMINAL/felleskat/lovbrudd.txt  
Utdatasett: \$KRIMINAL/anmeldt/arbeid/v030.ssd04  
Loggfil: \$KRIMINAL/anmeldt/logger/logg03.txt

- A. Hvis **lovbrudd** er forbrytelse ifølge lovbruddskatalogen settes **statbok** lik 'K'.
- B. Hvis **lovbrudd** er forseelse ifølge lovbruddskatalogen settes **statbok** lik 'F'.
- C. Hvis **lovbrudd** ikke er entydig forbrytelse eller forseelse ifølge lovbruddskatalogen beholdes **statbok** slik politiet oppgav den.



## 04 Omkodning av gjerningssted

Utgåtte kommunenumre for gjerningskommunen kodes om til gyldig kommunenummer etter kommunesammenslåing (Statistisk sentralbyrå 1994). Videre omkodes ugyldige kommunenumre for gjerningskommunen etter sone og politidistrikt. Ved manglende mulighet for å kode gjerningssted som kommune benyttes fylke der dette er mulig. Ferdig årgangsfil blir lagret under data/g[**aar**].ssd04

SAS-program: \$KRIMINAL/anmeldt/prog/bearbeid/p040.sas

Inndatsett: \$KRIMINAL/anmeldt/arbeid/v030.ssd04

Innfil: \$KRIMINAL/felleskat/polkam.txt  
 \$KRIMINAL/felleskat/kom\_gml.txt  
 \$KRIMINAL/felleskat/sonet.txt

Utdatsett: \$KRIMINAL/anmeldt/data/g[**aar**].ssd04

Loggfil: \$KRIMINAL/anmeldt/logger/log040.txt

- A. Enheter med **gjkom** ulik blank kontrolleres mot felles kommunekatalog. Enheter med ugyldige verdier og samtidig ulik spesielle kommunenummer (f.eks. Svalbard) gis ny verdi etter katalog over gamle kommuner (se vedlegg 7.2).
- B. Enheter med **gjkom** lik blank eller ugyldig gis ny kommuneverdi i **gjkom** etter **polkam** + **sonet** som anvist i sonekatalogen (vedlegg 7.2).
- C. Enheter med **gjkom** fremdeles blank eller ugyldig gis ny kommuneverdi i **gjkom** etter **polkam** som anvist i politikammerkatalogen (vedlegg 7.2) hvis den aktuelle kommunen er sammenfallende med politidistriktet eller utgjør mer enn 90 prosent av befolkningen i politidistriktet (gjelder per høsten 1999 ni politidistrikt).
- E. Enheter med **gjkom** fremdeles blank eller ugyldig gis *fylkesverdi* (dvs. ??00) i **gjkom** etter **polkam** + **sonet** som anvist i sonekatalogen (vedlegg 8.2).
- D. Enheter med **gjkom** ulik blank og ugyldig gis ny *fylkesverdi* (dvs. ??00) i **gjkom** etter de to første sifrene i **gjkom**, hvis de to første sifferene gir gyldig fylkesverdi eller er '21', '22', '23'.
- F. Enheter med ugyldig **gjkom** gis **gjkom** lik blank.

Tabell 2. Beskrivelse av det ferdig bearbejdede materialet, anmeldte lovbrudd

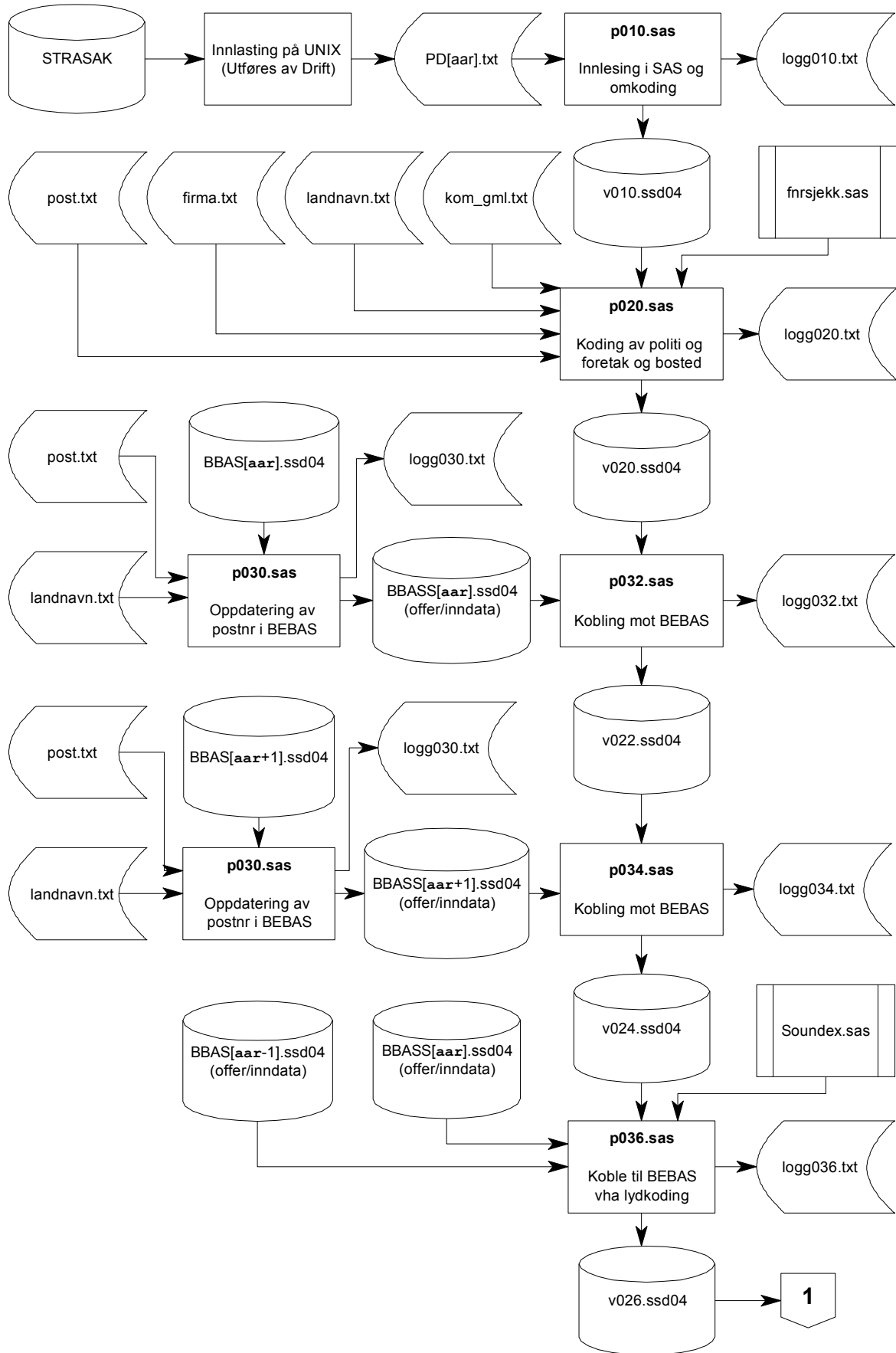
Kortnavn	Beskrivelse	Format	Gyldige verdier	Gyldighet
<b>Polkam</b>	Politikammer	char2	'01'-'55' , '57-62'	Alle år
<b>Aar</b>	Årstall	Num2		Alle år
<b>Saksnr</b>	Anmeldelsesnummer	char6	000001-	Alle år
<b>Saksbok</b>	Saksbokstav	Char2	Blank	Alle år
<b>Sekvens</b>	Sekvensnummer	Char4	Blank	Alle år
<b>Regdato</b>	Registreringsdato	Num6		Alle år
<b>Regaar</b>	Registreringsår	Num4	1990- <b>aar</b>	Alle år
<b>Regmnd</b>	Registreringsmåned	num2	1-12	Alle år
<b>Regdag</b>	Registreringsdag	num2		Alle år
<b>Lovbrudd</b>	Statistikkgruppe	char4		Alle år
<b>Modus</b>	Modus	Char4		Alle år
<b>Sone</b>	Sone for gjerningssted	Char4		92-
<b>Gjkom</b>	Gjerningsstedskommune	Char4		Alle år
<b>Gjdato1</b>	Gjerningsdato start	Num6		Alle år
<b>Gjaar1</b>	Gjerningsår start	Num4		Alle år
<b>Gjmnd1</b>	Gjerningsmåned start	num2		Alle år
<b>Gjdag1</b>	Gjerningsdag start	Num2		Alle år
<b>Gjdato2</b>	Gjerningsdato, slutt	Num6		Alle år
<b>Gjaar2</b>	Gjerningsår, slutt	Num4		Alle år
<b>Gjmnd2</b>	Gjerningsmåned slutt	num2		Alle år
<b>Gjdag2</b>	Gjerningsdag slutt	Num 2		Alle år
<b>Tidel</b>	Tidel, lovbruddsgruppe	char1		Alle år

### 3.3 Halvårsstatistikk over anmeldte lovbrudd

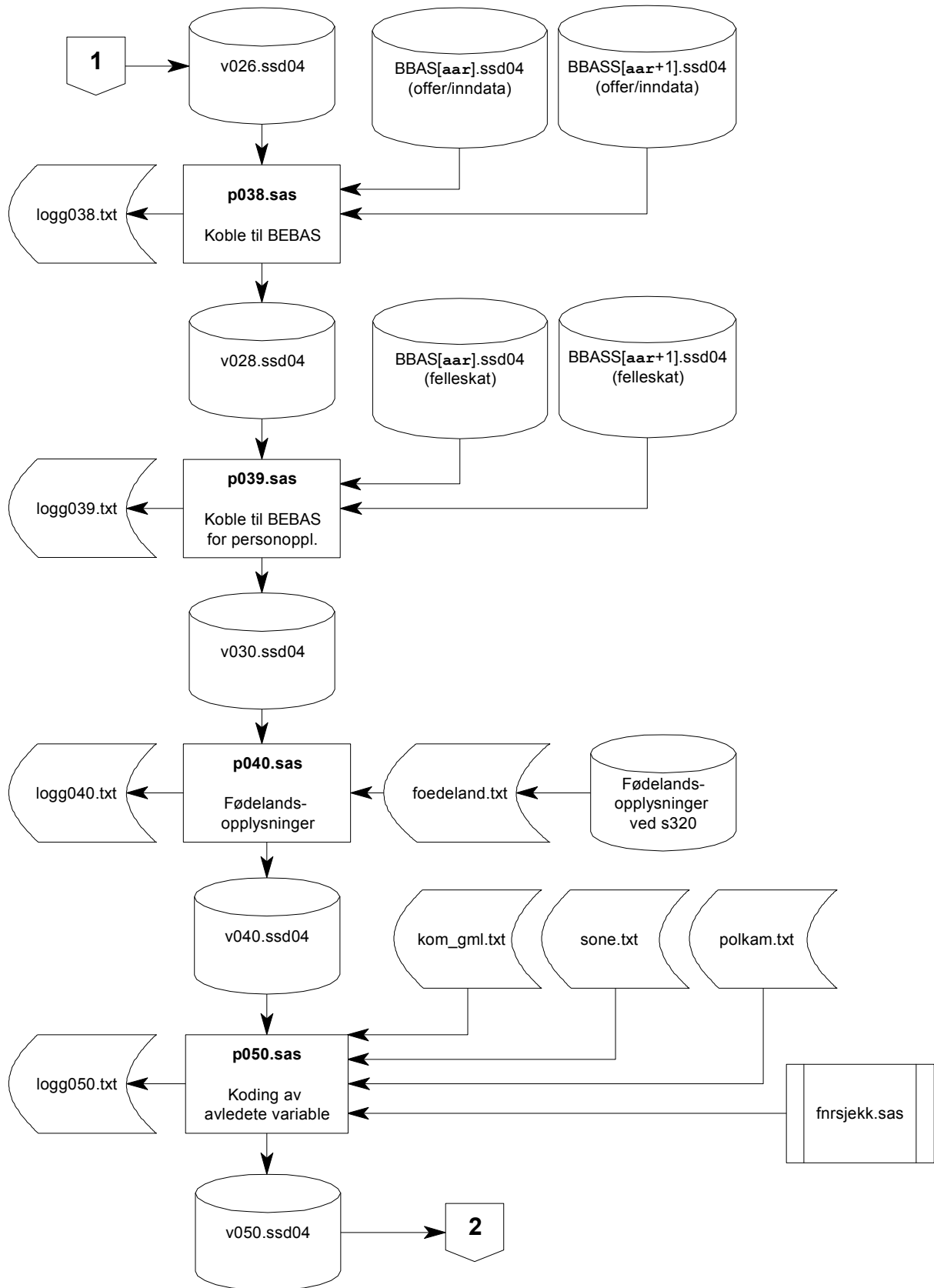
Andre uken i juli bestilles datamateriale for første halvår fra Politiets datatjeneste. Når halvårsstatistikk skal lages, følges samme bearbejding som i kapittel 3.2, men g[**aar**].txt og g[**aar**].ssd04 byttes ut med g[**aar**]h.txt og g[**aar**]h.ssd04. Rådatafilen gis PD[**aar**].txt som navn, men beholdes bare midlertidig. Filen bør slettes når materialet for hele statistikkåret kommer.

### 3.4 Registrerte ofre

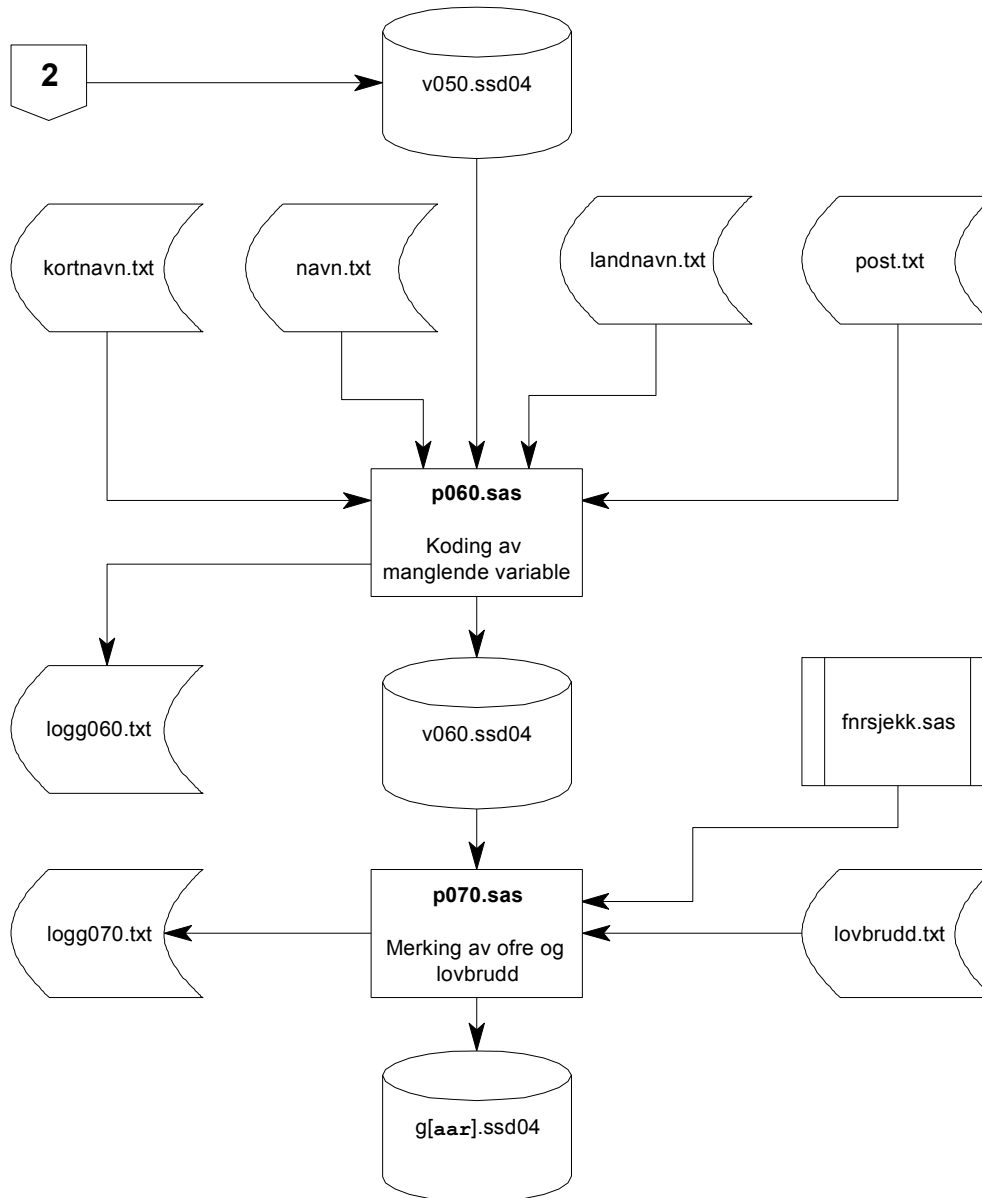
Produksjon av statistikk over registrerte ofre vil i praksis være å følge dataflyten i figur 4, figur 5 og figur 6, der alle navn på datasett, SAS-programmer og andre dokumenter korresponderer med de beskrevne rutinene i resten av dette kapitlet og i vedlegg 7.1 som viser filstrukturen på UNIX. Samme navn på SAS-programmer, datasett og logger benyttes for flere statistikker under kriminalstatistikken, så det henvises til enten filplassering i vedlegg 7.1 eller de enkelte rutinene nedenfor, for å entydig identifisere de enkelte delene i figur 4-6.



Figur 4. Logisk dataflyt, registrerte ofre, del 1



Figur 5. Logisk dataflyt, registrerte ofre, del 2



Figur 6. Logisk dataflyt, registrerte ofre, del 3

## 01 Innlesing fra rådata til SAS-datasett

SAS-program: \$KRIMINAL/offer/prog/bearbeid/p010.sas  
 Innfil: \$KRIMINAL/anmeldt/inndata/PD[aa].txt  
 Utdatasett: \$KRIMINAL/offer/arbeid/v010.ssd04  
 Loggfil: \$KRIMINAL/offer/logger/logg010.txt

- A. Enheter leses inn fra posisjoner på rådatafil formaterte variable i SAS-datasett, slik tabell X viser. Det må leses fra samme fil to ganger for å få med seg både saksopplysningene og personopplysningene.
- B. Alle enheter med **type** = 'KL' får oppdatert manglende kjennemerker (se forskjell mellom tabell 3 og 4) fra makkere blant enheter med **type** = 'SA' med hensyn til **polkam**, **saksnr**, **regaar**. Enheter med **type** = 'SA' uten makkere får **kjonn** = 'Q'.

Tabell 3. Filbeskrivelse av enhetene fra rådataene som viser saksopplysninger

Posisjon	Beskrivelse	Kortnavn	Format
1–2	Enhetstype ('SA' = saksopplysninger, 'KL' = Personoppysninger)	<b>Type</b>	char2
3–4	Politikammer	<b>Polkam</b>	char2
5–6	Registreringsår	<b>Aar</b>	num2
7–12	Saksnummer	<b>Saksnr</b>	char6
13–14	Delsaksbokstav	<b>Delsak</b>	char2
19–22	Registreringsår	<b>Regaar</b>	num4
23–24	Registreringsmnd	<b>Regmnd</b>	num2
25–26	Registreringsdag	<b>Regdag</b>	num2
27	Statistikkbokstav	<b>Statbok</b>	char1
28–31	Lovbruddskode fra STRASAK	<b>Lovbrudd</b>	char4
32–35	Modus	<b>Modus</b>	char4
36–39	Gjerningsstedssone	<b>Sone</b>	char4
40–43	Gjerningskommune	<b>Gjkom</b>	char4
44–47	Gjerningsår	<b>Gjaar</b>	num4
48–49	Gjerningsmåned	<b>Gjmnd</b>	num2
50–51	Gjerningsdag i måneden	<b>Gjdag</b>	num2

Tabell 4. Filbeskrivelse av enhetene fra rådataene som viser personopplysninger

Posisjon	Beskrivelse	Kortnavn	Format
1–2	Enhetstype ('SA' = saksopplysninger, 'KL' = Personoppysninger)	<b>Type</b>	char2
3–4	Politikammer	<b>Polkam</b>	char2
5–6	Registreringsår	<b>Aar</b>	num2
7–12	Saksnummer	<b>Saksnr</b>	char6
13–14	Delsaksbokstav	<b>Delsak</b>	char2
19–22	Fødselsår	<b>Faar</b>	num4
23–24	Fødselsmåned	<b>Fmnd</b>	num2
25–26	Fødselsdag	<b>Fdag</b>	num2
27–31	Personnummer	<b>Pnr</b>	char5
32–77	Slektsnavn	<b>Slnavn</b>	char46
78–123	Fornavn	<b>Fornavn</b>	char46
124–127	Postnummer	<b>Postnr</b>	char4
128–142	Poststed	<b>Postst</b>	char15
143	Status ('K' = offer, 'M' = melder, 'I' = implisert)	<b>Status</b>	char1

C. Det ryddes i variablene, dvs.:

- Hvis **postnr** = '0000', settes **postnr** lik blank
- Hvis **pnr** = '00000', settes **pnr** lik blank
- Hvis **sones** = '0000', settes **sones** lik blank
- Hvis **fmnd** = 0, settes **fmnd** lik blank
- Hvis **faar** = 0 og **fmnd** lik blank, settes **faar** lik blank
- Hvis **gjmnd** = 0, settes **fmnd** lik blank
- Hvis **gjaar** = 0 og **fmnd** lik blank, settes **faar** lik blank
- Hvis **regaar** - 210 < **faar** < **regaar** - 110, settes **faar** lik **faar** + 100
- Hvis **gjkom** = '0000', settes **gjkom** lik blank
- Hvis **gjkom** = '0000', settes **gjkom** lik blank

D. **Fdato** opprettes på grunnlag av **faar**, **fmnd** og **fdag**, og gis et SAS datoformat.

**Gjdato** opprettes på grunnlag av **gjaar**, **gjmnd** og **gjdag**, og gis et SAS datoformat.

**Regdato** opprettes på grunnlag av **regaar**, **regmnd** og **regdag**, og gis et SAS datoformat.

Variablene **type**, **fdag**, **gjdag** og **regdag** utelates i videre bearbeiding. Bare enheter med **type** = 'KL' eller **kjonn** = 'Q' beholdes.

- E. På grunn av at koblinger mot befolkningsdata, der **pnr** skal oppdateres, også baseres på en del andre felter må det ryddes i disse for å få mest mulig korrekte personopplysninger, der det er mulig:
- Alle doble blanke byttes ut med enkle blanke i **fornavn**, **slnavn** og i **postst**
  - '-' byttes ut med '.' i **fornavn** og i **slnavn**
  - '-' byttes ut med '.' i **fornavn** og i **slnavn**
  - Kjente forkortelser for navn byttes ut med hele navnet.
  - Tekststrengene 'JR.', 'JR', 'J.R.', 'JR ', 'SN.', 'SR.', 'D.Y.', 'D.E.' i **fornavn** og **slnavn** fjernes<sup>2</sup> og **kjonn** settes lik '1' hvis forkortelsene finnes i enten **slnavn** eller **fornavn**.
  - Hvis **fdato** er større enn **regdato** settes **fdato**, **fmnd** og **faar** lik blank  
Hvis **faar** > **regaar**, eller **faar** = **regaar** og **fmnd** > **regmnd**, settes **faar** og **fmnd** lik blank. Hvis **fdato** og **pnr** ikke gir gyldig fødselsnummer settes **pnr** lik blank.
- F. Det gjennomføres en fullstendig dublettkontroll. Det beholdes bare en enhet blant enheter med samme verdi i **polkam**, **aar**, **saksnr**, **delsak**, **lovbrudd**, **slnavn**, **fornavn**, **postnr**, **postst**, **status**, **fdato** og **pnr**.
- G. Enheter med **lovbrudd** > '9199' eller **polkam** lik '56' eller **statbok** ulik 'F' og 'K' slettes.

## 02 Koding av enheter fra politi og foretak og koding av bosted

SAS-program: \$KRIMINAL/offer/prog/bearbeid/p020.sas  
 Inndatsett: \$KRIMINAL/offer/arbeid/v010.ssd04  
 Innfil: \$FELLES/sasprog/fnr sjekk.sas  
           \$KRIMINAL/offer/kat/firma.txt  
           \$KRIMINAL/felleskat/landnavn.txt  
           \$KRIMINAL/felleskat/post.txt  
 Utdatsett: \$KRIMINAL/offer/arbeid/v020.ssd04  
 Loggfil: \$KRIMINAL/offer/logger/logg020.txt

- A. For enheter med **pnr** = '99500' settes **kjonn** = 'F'.
- B. Det letes i **slnavn** og **fornavn** etter tekststrenger som indikerer at anmeldelsen kommer fra politiet. Ved treff settes **kjonn** lik 'P'.
- C. For enheter med **kjonn** lik blank letes det i **slnavn** og **fornavn** etter tekststrenger som indikerer foretak. Ved treff settes **kjonn** lik 'F'.
- D. For enheter med **kjonn** lik 'P', 'Q' eller 'F' settes **pnr** = blank<sup>3</sup>. For enheter med **kjonn** ulik 'P', 'Q' eller 'F', og ugyldig fødselsnummer settes **pnr** = blank.

2 Noen av disse forkortelsene kan representere virkelige navn, men siden de er forkortelser vil de uansett ikke oppnå treff i befolkningsdatabasen. Hvis enheten har gyldig fødselsnummer (nå eller innen rutine 3, del 6) vil **kjonn** overskrives senere.

3 Blanking av **pnr** gjøres for å forhindre at et gyldig fødselsnummer skal overstyre funn av foretak i tekstfeltene. Erfaringsmessig har det vist seg at flere foretak blir registrert med fødselsnummeret til den personen som anmelder lovbruddet på vegne av foretaket. Det er dermed en viss risiko for å stryke en person hvis man feilaktig finner en tekst i navnet som tolkes som foretak, men ved å beholde **fdato** holdes muligheten åpen for at **slnavn**, **fornavn**, **fdato**, **bokom** og **postnr** skal overstyre 'F' i **kjonn** gjennom koblinger i rutine 03.

Datasettet oppdateres så langt det er mulig med bostedsopplysninger før koblingene til BEBAS starter. Det er viktig å tilordne **bokom** og **postnr** til rette verdier for senere kobling, men det er også viktig å blanke ut gale verdier, slik at disse ikke står i veien for senere kobling mot BEBAS.

- E. Variabelen **bokom** opprettes. Hvis **postnr** er gyldig får **bokom** verdi etter katalog (se postnummerkatalog, vedlegg 7.2). Hvis **bokom** fortsatt er blank og **postst** ikke er blank får **bokom** verdi etter hele **postst**, det vil si at det i første omgang ikke godtas at bare deler av **postst** utgjør et poststedsnavn (se postnummerkatalog, vedlegg 7.2). **Bokom** testes og eventuelt omkodes etter gamle kommunenumre (se katalog over utgåtte kommunenumre, vedlegg 7.2).
- F. Variabelen **boland** opprettes. Hvis **postnr** lik '0001', '0002' eller '0003' settes **boland** til henholdsvis '106', '101' og '102' og **postnr** settes lik blank.  
Ved blank **boland** og **bokom**, og **postst** ulik blank, letes det etter landnavn i **postst**. Hvis treff settes **boland** til korresponderende landkode og **postnr** settes blank.  
Ved blank **bokom** og **boland**, og **postst** ulik blank, letes det etter tall med fem siffer i **postst**. Ved funn settes **boland** lik '950' og **postnr** lik blank.  
Med fremdeles blank **bokom**, **boland** lik blank eller '950', og **postst** ulik blank, letes det etter postnummer, kommunenavn og poststedsnavn, der poststedsnavn og kommunenavn er på mer enn tre bokstaver, i **postst**. Ved funn settes **bokom** til korresponderende kommunenummer og **postnr** lik blank.

### 03 Tilføyelse av personopplysninger fra BEBAS

På grunn av tunge SAS-programmer som i praksis bør kjøres om natten, er denne rutinen delt opp i flere deler, der datasettet lagres i katalog \$KRIMINAL/anmeldt/arbeid mellom hver kjøring, bortsett fra uttrekket av BEBAS (Del 1), som lagres under \$KRIMINAL/offer/inndata.

For rutinene p030.sas til p039.sas må saksbehandler oppgi hvilken årgang som skal benyttes som inndata i rutinen. Dette gjøres ved at det står en eller flere linjer øverst i programmet (for eksempel %LET &aaret = 1998) som saksbehandler kan skriv inn årgangen(e) i, og som resten av programmet bruker når det henter datafilene for kobling.

#### *Del 1:*

SAS-program: \$KRIMINAL/offer/prog/bearbeid/p030.sas  
Inndatasett: \$KRIMINAL/felleskat/BBAS[**aar**].ssd04  
Innfiler: \$KRIMINAL/felleskat/post.txt  
\$KRIMINAL/felleskat/landnavn.txt  
Utdatasett: \$KRIMINAL/offer/inndata/BBAS[**aar**].ssd04  
Loggfil: \$KRIMINAL/anmeldt/logger/logg030.txt

Det skal ligge en versjon av Det sentrale folkeregisteret/BEBAS på fellesområdet for kriminalstatistikk som gjelder 1. januar statistikkåret. Her skal også en versjon for 1. januar året etter statistikkåret (dvs. **aar**+1) ligge. Hvis disse datasettene er opprettet er denne rutinen unødvendig å gjennomføre for å fortsette bearbeidingen. Datasettene benyttes for flere statistikkområder under kriminalstatistikk og opprettelse og sletting må samkjøres med disse. Hvis kun ascii-filer leveres av seksjon 320 eller seksjon 303, eksisterer det et program for innlesning



som oppretter de aktuelle variablene, og som kan tilpasses den leverte ascii-filen (\$KRIMINAL/eforsk/prog/bbasINN.sas).

Rutinen kjøres på nytt for hver årgang av BEBAS som er nødvendig for å gjennomføre rutinene p032.sas til p038.sas.

Først oppdateres **postnr** etter de tre andre adressefeltene i BEBAS-datasettene, der **postnr** er blank.

A. Hvis **postnr** lik blank letes det i **postadr1**, **postadr2** og **postadr3** etter gyldige postnumre. Hvis gyldig postnummer finnes settes **postnr** lik dette. Unntakene er der det kan antas at adressen refererer til utlandet eller nummeret referer til andre størrelser enn postnummer. Utenlandsadresser identifiseres ved å lete i de samme feltene etter navn på **land**. (katalog over landnavn, se vedlegg 7.2), og ved å teste på om sifferet som kan være et postnummer er del av et tall med flere enn fire siffer (mange andre land har postnummer med fem siffer). Det testes også på om nummeret indikerer gatenummer, romnummer o.l. Hvis landnavn finnes eller tallet er på fem siffer eller mer eller det finnes tekst som indikerer at sifferet kommer fra noe annet enn postnummer, oppdateres ikke **postnr**.

B. **Bokom** omkodes for gamle og ugyldige kommunenumre, se katalog i vedlegg 7.2

#### Del 2:

SAS-program: \$KRIMINAL/offer/prog/bearbeid/p032.sas  
 Inndatasett: \$KRIMINAL/offer/arbeid/v020.ssd04  
                   \$KRIMINAL/offer/inndata/BBAS[**aar**].ssd04  
 Innfiler: \$FELLES/ssbkat/poststed/post.txt  
 Utdatasett: \$KRIMINAL/offer/arbeid/v022.ssd04  
 Loggfil: \$KRIMINAL/anmeldt/logger/logg032.txt

Datasettet oppdateres så langt det er kvalitetsmessig forsvarlig med personnummer etter andre variabelverdier som kan identifisere personen, gjennom en lang rekke koblinger til BEBAS. Det tas utgangspunkt i **fdato**, **slnavn**, **fornavn**, **postnr** og **bokom** som variable som kan identifisere personene. Det vil i prinsippet da kunne inntreffe at en enhet kan bli koblet til feil person, hvis det er flere personer med identiske kjennemerker.

I koblingene er det satt en grense ved fire prosent for sannsynlig feilkobling. Den rutinen som kommer nærmest denne grensen er LGF og N. Disse har en sannsynlighet for feilkobling på ca 3,3 og 3,8 prosent<sup>4</sup>, og er blant annet av den grunn plassert sist i rekken av koblinger. Ingen av de andre rutinene har en feilmargen på mer enn 1,4 prosent. En oversikt over rutinene, blant annet over risikoen for feilkobling, er satt opp i tabell 5.

4 Sannsynligheten for feilkobling er anslått på grunnlag av antall dubletter i BEBAS (koblingen skjer til den første). Selv med dubletter betyr det ikke at man kobler feil, det kan tilfeldigvis være den riktige enheten som det kobles til. Sannsynligheten for dette er avhengig av antall dubletter: flere dubletter gir mindre sannsynlighet for å treffe riktig. En annen egenskap ved kobling tilsier at man ikke skal være så redd for feilkobling: kobling til dublett vil ofte gi samme resultat for variabelverdiene som det en kobling til den rette enheten gir. Eksempelvis vil fornavn lik 'Håkon', lik fødselsdato, postnummer og etternavn sannsynligvis gi samme verdi for bokommune, kjønn, statsborgerskap og landbakgrunn, selv om man kobler til en dublett.

Selve koblingene startes, der punktene C til F gjentas for forskjellige kombinasjoner av personopplysninger og årganger:

- C. Enheter med ugyldig fødselsnummer og med makkere i BBAS[**aar**].ssd04 med hensyn til **slnavn**, **fornavn**, **postnr** og **fdato**, der alle nevnte variable er ulik blank, får oppdatert **pnr**.
- D. Enheter med ugyldig fødselsnummer og med makkere i BBAS[**aar**].ssd04 med hensyn til **slnavn**, **fornavn**, **fdato** og **bokom**, der alle nevnte variable er ulik blank, får oppdatert **pnr**.
- E. Enheter med ugyldig fødselsnummer og med makkere i BBAS[**aar**].ssd04 med hensyn til **slnavn**, **fornavn** og **fdato**, der alle nevnte variable er ulik blank, og samtidig **postnr** lik blank eller **postnr** fra BEBAS lik blank, får oppdatert **pnr**.
- F. Enheter med **fdato** og **pnr** lik blank og med makkere i BBAS[**aar**].ssd04 med hensyn til **slnavn**, **fornavn** og **postnr** der alle nevnte variable er ulik blank får oppdatert **pnr** og **fdato**.

Siden registrering av fornavn hos politiet ikke er underlagt kvalitetskontroller i samme grad som i BEBAS, antas det at et eventuelt annet fornavn kan ha blitt utelatt ved registrering hos politiet, eller at det også registreres et mellomnavn i feltet for fornavn. Det kobles derfor med kun første ord i **fornavn** i BEBAS mot første ord i **fornavn** i datamaterialet fra politiet.

- G. Rutinene C til F gjentas med å benytte kun første ord i **fornavn**.
- H. Variabelen **slnavn2** benyttes som **slnavn**. og koblingene C–G utføres på nytt.
- I. Enheter med manglende **pnr** og med makkere i BBAS[**aar**].ssd04 med hensyn til **slnavn**, **postnr** og **fdato**, der alle nevnte variable er ulik blank, og samtidig **fornavn** lik blank, får oppdatert **pnr**.
- J. Enheter med manglende **pnr** og med makkere i BBAS[**aar**].ssd04 med hensyn til **slnavn2**, **postnr** og **fdato**, der alle nevnte variable er ulik blank, og samtidig **fornavn** lik blank, får oppdatert **pnr**.

*Del 3:*

SAS-program: \$KRIMINAL/offer/prog/bearbeid/p034.sas  
Inndatsett: \$KRIMINAL/offer/arbeid/v022.ssd04  
\$KRIMINAL/offer/inndata/BBAS[**aar**+1].ssd04  
Utdatsett: \$KRIMINAL/offer/arbeid/v024.ssd04  
Loggfil: \$KRIMINAL/offer/logger/log034.txt

- K. Punktene C til J gjentas med BBAS[**aar**+1].

*Del 4:*

SAS-program: \$KRIMINAL/offer/prog/bearbeid/p036.sas  
 Inndatsett: \$KRIMINAL/offer/arbeid/v024.ssd04  
               \$KRIMINAL/offer/inndata/BBAS[**aar**].ssd04  
               \$KRIMINAL/offer/inndata/BBAS[**aar+1**].ssd04  
 Innfiler: \$KRIMINAL/offer/prog/Soundex.sas  
 Utdatsett: \$KRIMINAL/offer/arbeid/v026.ssd04  
 Loggfil: \$KRIMINAL/offer/logger/logg036.txt

Det er utviklet en rutine som koder om tekststrenger til "rene" tekststrenger, dvs. mest mulig nære lydskrift. Dette er imidlertid en svært "tung" rutine å kjøre, med dagens maskinvare tar den minimum 8 timer å eksekvere, så koblingsforsøkene gjentar derfor ikke alle tidligere forsøk. Koblingene begrenses derfor til de som gir størst gevinst i form av antall treff. Ved fremtidige oppgraderinger av UNIX-tjenerene bør det vurderes om alle punktene fra C til J bør kjøres med omkodet **fornavn** og **slnavn**.

L. Punkt GC til GF kjøres på nytt med omkodet **slnavn** og første del av **fornavn** til lydskrift.

M. Punkt L gjentas med BBAS[**aar+1**].ssd04, dvs. punkt KGC til KGF med omkodet **slnavn** og første del av **fornavn**.

*Del 5:*

SAS-program: \$KRIMINAL/offer/prog/bearbeid/p038.sas  
 Inndatsett: \$KRIMINAL/offer/arbeid/v026.ssd04  
               \$KRIMINAL/offer/inndata/BBAS[**aar**].ssd04  
               \$KRIMINAL/offer/inndata/BBAS[**aar+1**].ssd04  
 Utdatsett: \$KRIMINAL/offer/arbeid/v028.ssd04  
 Loggfil: \$KRIMINAL/offer/logger/logg038.txt

For de enhetene med både **pnr** og **fdato** lik blank ble det tidligere forsøkt å koble ved hjelp av **postnr** (rutine F). **Postnr** er imidlertid i mindre grad oppgitt i BEBAS, mens **bokom** er oppgitt i neste alle enhetene. Bokommune er imidlertid en så stor geografisk enhet at vi må legge begrensninger på **fornavn** for å få risikoen for feilkobling under fire prosent. Med krav om at **fornavn** skal være på minst fem bokstaver oppfylles kravet til sikkerhet mot feilkobling.

N. Enheter med **fdato** og **pnr** lik blank og med makkere i BBAS[**aar**].ssd04 med hensyn til **slnavn**, **fornavn** og **bokom** der alle nevnte variable er ulik blank, og **fornavn** er på minst fem bokstaver, får oppdatert **pnr** og **fdato**. Koblingen gjentas med BBAS[**aar+1**].ssd04.

Variabelen **bokom** er god oppgitt i BEBAS-materialet, men ikke i like stor grad i offermaterialet. Det gjennomføres derfor en kobling på de enhetene med blank **fdato** og **pnr** der **gjkom** benyttes der **bokom** er blank. For å sikre seg mot mulige feilkoblinger legges det restriksjoner på **fornavn** tilsvarende som i rutine N.

O. Enheter med **fdato** og **pnr** lik blank og med makkere i BBAS[**aar**].ssd04 med hensyn til **slnavn**, **fornavn** og **gjkom** som **bokom** der alle nevnte variable er ulik blank, **fornavn** er på minst fem bokstaver og **bokom** er blank, får oppdatert **pnr** og **fdato**. Koblingen gjentas med BBAS[**aar+1**].ssd04.

Etter koblingene skal 50 forskjellige varianter for identifisering av personer uten **pnr** være utprøvd. En oppsummering finnes i tabell 5 og resultatene kan delvis ses av figur 2, side 7. Alle koblinger kjøres for både **aar** og **aar+1**, men risiko for feilkobling som kobles er beregnet for førstegangs kobling, dvs. **aar**. Gjennomføringen av rutinene for statistikkåret 1998 viser at litt over 64 prosent av personene med manglende **pnr** fikk påført **pnr** ved kobling mot BBAS1998.ssd04, mens litt under 2 prosent fikk påført **pnr** ved kobling til BBAS1999.ssd04.<sup>5</sup> En god del av resten av enhetene med uoppgitt **pnr** vil være personer fra andre land som har vært utsatt for en kriminell handling i Norge. Disse merkes gjennom rutine p020-F og p060-A. Andelen av registrerte ofre som ved bearbeidingsens slutt gjenstod med uoppgitt **pnr** og var dermed for 1998 på 3,1 prosent.

**Tabell 5. Koblingene mot BEBAS, kjennemerker og risiko for feilkobling**

Rutine	Variabelnavn							Risiko for feilkobling (%)
	Slnavn	Slnavn2	Fornavn 1.	fornavn	Fdato	Postnr	Bokom	
C	G		G		G	G		< 0,01
D	G		G		G		G	< 0,01
E	G		G		G			0,02
F	G		G		B	G		0,35
GC	G			G	G	G		< 0,01
GD	G			G	G		G	< 0,01
GE	G			G	G			0,04
GF	G			G	B	G		1,4
HC		G	G		G	G		< 0,01
HD		G	G		G		G	< 0,01
HE		G	G		G			< 0,01
HF		G	G		B	G		0,2
HGC		G		G	G	G		< 0,01
HGD		G		G	G		G	< 0,01
HGE		G		G	G			0,02
HGF		G		G	B	G		0,65
I	G		B		G	G		0,25
J		G	B		G	G		0,2
LGC	L			L	G	G		< 0,01
LGD	L			L	G		G	< 0,01
LGE	L			L	G			0,11
LGF	L			L	B	G		3,3
N	G		G*		B		G	3,8
O	G		G*		B		G**	3,2

Der det står en bokstav i tabellruten kreves det treff mellom BEBAS og offermaterialet. Der ruten er blank sjekkes det ikke på overensstemmelse.

G = Krav om gyldig verdi i BEBAS og i offermaterialet.

B = Krav om blank verdi i offermaterialet.

L = Krav om lik lydkode/lydskrift på variablene i BEBAS og i offermaterialet.

\* = Krav om **fornavn** på minst 5 bokstaver.

\*\* = Kobler bare hvis **bokom** er blank i offermaterialet og **gjkom** i offermaterialet er lik **bokom** i BEBAS.

5 Grunnen til at så få enheter får påført **pnr** fra BBAS[**aar+1**].ssd04 er først og fremst at alle koblinger prøves først mot BBAS[**aar**].ssd04.

*Del 6:*

SAS-program: \$KRIMINAL/offer/prog/bearbeid/p039.sas  
 Inndatasett: \$KRIMINAL/offer/arbeid/v028.ssd04  
                   \$KRIMINAL/felles/BBAS[**aar**].ssd04  
                   \$KRIMINAL/felles/BBAS[**aar+1**].ssd04  
 Innfil: \$FELLES/sasprog/fnrsjekk.sas  
 Utdatasett: \$KRIMINAL/offer/arbeid/v030.ssd04  
 Loggfil: \$KRIMINAL/offer/logger/logg039.txt

- N. Enheter med makkere med hensyn til **fdato**, **slnavn** og **fornavn**, der alle variablene er ulik blank, får oppdatert **pnr** fra makker, hvis **pnr** er blank og makkerens fødselsnummer er gyldig.
- O. Enheter med gyldige fødselsnummer tilføres data fra BEBAS med data som gjaldt 1. januar i statistikkåret (BBAS[**aar**].ssd04) og data som gjaldt 1. januar i året etter statistikkåret (BBAS[**aar+1**].ssd04). De aktuelle variablene i BBAS[**aar**] og BBAS[**aar+1**] døpes om med endelsen 'jan' og 'des' (i forhold til statistikkåret) før kobling for å teste på blanke variable.
- Hvis **bokomjan** ulik blank settes **bokom** lik **bokomjan**. Hvis **bokomjan** er blank og **bokomdes** ulik blank settes **bokom** lik **bokomdes**.
  - Hvis **statjan** ulik blank settes **stat** lik **statjan**. Hvis **statjan** er blank og **statdes** ulik blank settes **stat** lik **statdes**.

**Tabell 6. Filbeskrivelse av BBAS[**aar**].ssd04 og BBAS[**aar+1**].ssd04**

Kortnavn	Beskrivelse	Format
<b>fdato</b>	Fødselsdato	ddmmyy
<b>pnr</b>	Personnummer	char5
<b>slnavn</b>	Slektsnavn	char50
<b>fornavn</b>	Fornavn	char50
<b>slnavn2</b>	Alternativt slektsnavn	char50
<b>postadr1</b>	Bostedsadresse	char30
<b>postadr2</b>	Alternativ bostedsadresse	char30
<b>postadr3</b>	Alternativ bostedsadresse	char30
<b>postnr</b>	Postnummer for bostedsadresse	char4
<b>postst</b>	Poststed for bostedsadresse	char30
<b>bokom</b>	Bokommune 1. januar	char4
<b>stat</b>	Statsborgerskap 1. januar	char3

#### 04 Koding av landbakgrunn ved hjelp av "fødelandsfil"

SAS-program: \$KRIMINAL/offer/prog/bearbeid/p040.sas  
 Inndatasett: \$KRIMINAL/offer/arbeid/v030.ssd04  
 Innfil: \$KRIMINAL/fodeland/g[**aar**].txt  
 Utdatasett: \$KRIMINAL/offer/arbeid/v040.ssd04  
 Loggfil: \$KRIMINAL/offer/logger/log040.txt

I likhet med programmene under punkt 03, må saksbehandler også i denne rutinen oppgi hvilken årgang som skal benyttes av inndata.

For hver person skal det hentes opplysning om innvandrersstatus, for eksempel landbakgrunn og type innvandrers. Disse opplysningene finnes på en "fødelandsfil" som utarbeides av seksjon 320, og som kopieres til kriminalstatistikken sitt hjemmeområde. Det må benyttes en fødelandsfil som

er situasjonsfil 1. januar året etter statistikkåret for å få med eventuelle innvandrere som ble registrert i statistikkåret.

- A. Alle enheter får påført **botid**, **bydel**, **invkat**, **ib**, **ibland**, **ibvereg** og **gen** fra fødelandsfilen etter kobling med **fdato** og **pnr** som nøkkel.

**Tabell 7. Filbeskrivelse av \$KRIMINAL/fodeland/g[aar].txt<sup>6</sup>**

Posisjon	Beskrivelse	Kortnavn	Format
1–11	Fødselsnummer	<b>fnr</b>	char11
49–50	Botid i hele år	<b>botid</b>	num2
51–54	Bokommune	<b>bokom</b>	char4
59–60	Bydel i Oslo	<b>bydel</b>	char2
73	Innvandrerkategori	<b>invkat</b>	char1
78	Innvandrerstatus	<b>ib</b>	char1
79–81	Bakgrunnsland	<b>ibland</b>	char3
82	Verdensregion, bakgrunnsland	<b>ibvereg</b>	char1
88	Generasjon	<b>gen</b>	char1

- B. For de enhetene med **bokom** lik blank og gyldig fødselsnummer hentes **bokom** fra fødelandsfilen.

## 05 Koding av avledede variable

SAS-program: \$KRIMINAL/offer/prog/bearbeid/p050.sas

Inndatasett: \$KRIMINAL/offer/arbeid/v040.ssd04

Innfil: \$KRIMINAL/felleskat/kom\_gml.txt

\$KRIMINAL/felleskat/polkam.txt

\$KRIMINAL/felleskat/sone.txt

\$FELLES/sasprog/fnrsjekk.sas

Utdatasett: \$KRIMINAL/offer/arbeid/v050.ssd04

Loggfil: \$KRIMINAL/offer/logger/logg050.txt

Gjerningskommune testes og eventuelt omkodes:

- A. Enheter med **gjkom** ulik blank kontrolleres mot felles kommunekatalog. Enheter med ugyldige verdier og samtidig ulik spesielle kommunenummer (f.eks. Svalbard) gis ny verdi etter katalog over gamle kommuner (se vedlegg 7.2).
- B. Enheter med **gjkom** lik blank eller ugyldig gis ny kommuneverdi i **gjkom** etter **polkam** + **sone** som anvist i sonekatalogen (vedlegg 7.2).
- C. Enheter med **gjkom** fremdeles blank eller ugyldig gis ny kommuneverdi i **gjkom** etter **polkam** som anvist i politikammerkatalogen (vedlegg 7.2) hvis den aktuelle kommunen er sammenfallende med politidistriktet eller utgjør mer enn 90 prosent av befolkningen i politidistriktet (gjelder per høsten 1999 ni politidistrikt).
- D. **gjfylke** gis verdi etter de to første sifrene i **gjkom**, der **gjkom** er en gyldig kommunekode eller begynner på '21', '22', '23'.

6 Tabellen dekker bare de kjennemerkene som er aktuelle for kriminalstatistikken. For en komplett filbeskrivelse henvises det til seksjon 320.

- E. Enheter med **gjfylke** lik blank gis **gjfylke** etter **polkam** + **son**e som anvist i sonekatalogen (vedlegg 8.2).
- F. Enheter med **gjfylke** lik blank og **gjkom** ulik blank, men ugyldig, gis **gjfylke** etter de to første sifrene i **gjkom**, hvis disse gir gyldig fylkeskode. **Gjkom** settes deretter lik blank.
- G. Enheter med **gjfylke** fremdeles lik blank gis fylkesverdi etter **polkam** som anvist i politikammerkatalogen (vedlegg 7.2).

Bokommune testes og eventuelt kodes om. Dette er den samme rutinen som gjøres under p032-A, men foretas nå også på data fra BEBAS og fødelandsfilen:

- H. Enheter med **bokom** ulik blank kontrolleres mot felles kommunekatalog. Enheter med ugyldige verdier og samtidig ulik spesielle kommunenummer (f.eks. Svalbard) gis ny verdi etter katalog over gamle kommuner (se vedlegg 7.2).

Bokommunen oppdateres etter makkerer:

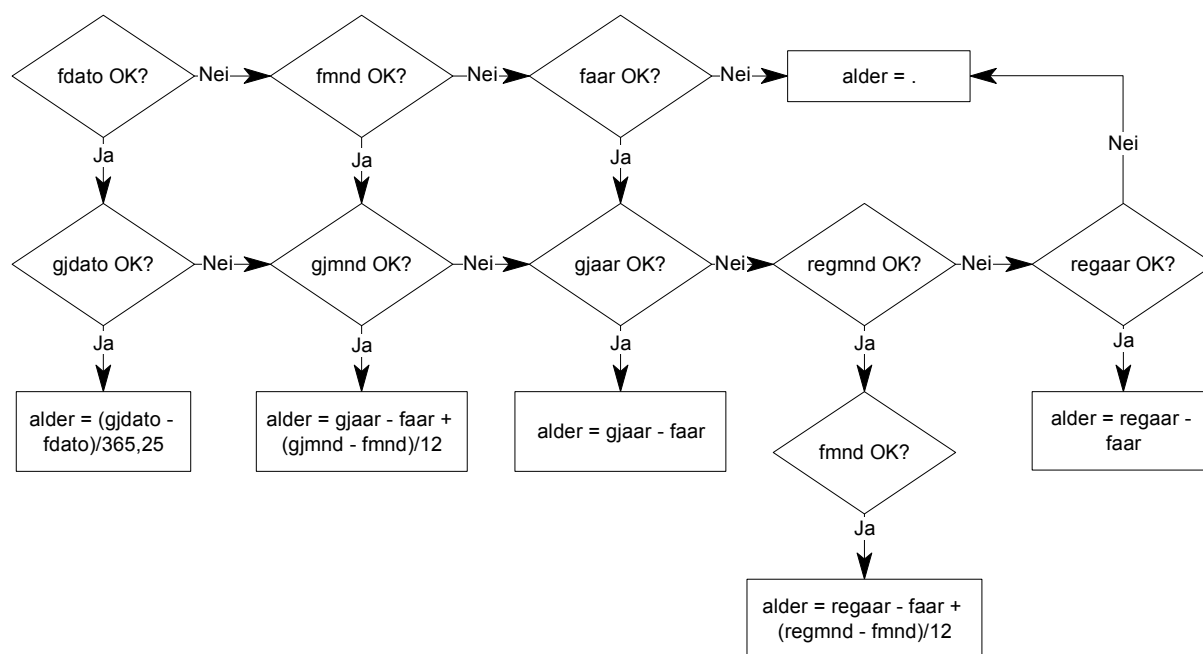
- I. Enheter med gyldig fødselsnummer og med blank **bokom** får tilført verdier i **bokom** hvis noen av makkerne med hensyn til fødselsnummeret har **bokom** ulik blank. **bokom** settes da lik makkerens **bokom**.

Alder beregnes for de enhetene som representerer personer fra tilgjengelige datovariabler. Utrekningen av alder skjer i hele år, dvs. antall fylte år på gjerningstidspunktet.

- J. For enheter med **fdato** større enn **regdato** settes **fdato** lik blank. For enheter med **gjdato** større enn **regdato** eller **gjdato** mindre enn **fdato** settes **gjdato**, **gjmn**d og **gjjaar** lik blank.
- K. For enheter med **kjonn** ulik 'P', 'Q' og 'F': Hvis **gjdato** og **fdato** ikke er blanke: **alder** =  $(\text{gjdato} - \text{fdato})/365,25$ . Hvis enten **gjdato** eller **fdato** er blank benyttes **fmnd/faar** og **gjmn**d/**gjjaar**, hvis disse ikke er blanke: **alder** =  $(\text{gjjaar} - \text{faar}) + (\text{gjmn}d - \text{fmnd})/12$ . Hvis enten **fmnd** eller **gjmn**d er blank benyttes årsdelen: **alder** = **gjjaar** - **faar**. Dersom **gjjaar** er blank og **fmnd**, **faar**, **regaar** og **regmn**d ikke er blanke: **alder** =  $(\text{regaar} - \text{faar}) + (\text{regmn}d - \text{fmnd})/12$ . Hvis **fmnd** er blank: **alder** = **regaar** - **faar**. Hvis ethvert grunnlag for beregning av alder mangler, gis **alder** verdi blank.

Registreringsdatoen skal ikke benyttes hvis det er snakk om seksuelle overgrep mot mindreårige, det vil si i lovbrudd der offerstatusen har sammenheng med alder.

- K. Hvis **alder** > 13, **lovbrudd** lik '1402' eller '1403' og **gjjaar** lik blank, settes **alder** lik blank. Hvis **alder** > 15, **lovbrudd** lik '1404' eller '1418' og **gjjaar** lik blank, settes **alder** lik blank.



Figur 7. Beregning av alder

Det opprettes en variabel (**type**) som kodes etter lovbruddets art<sup>7</sup>:

- L. Enheter som etter **lovbrudd** per definisjon er forbrytelser gis **type** = '1' ;  
 Enheter som etter **lovbrudd** per definisjon er forseelser gis **type** = '0' ;  
 Enheter som ikke er entydig forbrytelse eller forseelse etter **lovbrudd** gis **type** etter

**bokstav**:

- Hvis **bokstav** = 'K' settes **type** lik '1'.
- Hvis **bokstav** = 'F' settes **type** lik '0'.

Personens kjønn kodes etter tredje siffer i **pnr**:

- M. Hvis **kjonn** lik blank eller fødselsnummeret er gyldig, og tredje siffer i **pnr** er '0', '2', '4', '6' eller '8' settes **kjonn** lik '2'.
- Hvis **kjonn** lik blank eller fødselsnummeret er gyldig, og tredje siffer i **pnr** er '1', '3', '5', '7' eller '9' settes **kjonn** lik '1'.

<sup>7</sup> Dette skulle i utgangspunktet vært gjort i STRASAK, men Politiets datatjeneste har opplyst at manglende oppdatering forekommer.



## 06 Koding av variabelverdier

SAS-program: \$KRIMINAL/offer/prog/bearbeid/p060.sas  
 Inndatsett: \$KRIMINAL/offer/arbeid/v050.ssd04  
 Innfil: \$KRIMINAL/offer/kat/navn.txt  
 \$KRIMINAL/offer/kat/kortnavn.txt  
 \$KRIMINAL/felleskat/landnavn.txt  
 \$KRIMINAL/felleskat/post.txt  
 \$FELLES/sasprog/fnr sjekk.sas  
 Utdatsett: \$KRIMINAL/offer/arbeid/v060.ssd04  
 Loggfil: \$KRIMINAL/offer/logger/logg060.txt

Det testes om personene med **bokom** lik blank bor i utlandet

- A. Hvis **postnr** lik blank eller **bokom** lik blank og ugyldig fødselsnummer<sup>8</sup> letes det etter landnavn og tekststrenger i **poststed** som indikerer utlandet. Hvis slik tekst finnes settes **boland** lik eventuelt landkode.
- B. Hvis **bokom** ulik blank og **boland** lik blank og **kjonn** lik '1', '2' eller blank, settes **boland** lik '000'.

For de enhetene med **kjonn** lik blank letes det etter egennavn i **fornavn** som bestemmer kjønn. Som grunnlag er det opprettet en fil over alle fornavn som er gitt barn i Norge i 1997. Denne filen er satt sammen av både gutt- og jentenavn. Også oppsplittede dobbelnavn er inkludert. En eventuell oppdatering av materialet vil bære preg av utviklingsarbeid, både fordi en må ta stilling til om nåværende materiale skal erstattes eller suppleres, hvordan navn som er benyttet for begge kjønn skal behandles og så videre, og fordi rådata fra navnestatistikken ikke er på en fastlagt og formalisert form. Produksjon av dette materialet er derfor ikke dokumentert.

Der samme navnet er benyttet for begge kjønn brukes navnet hvis ett av kjønnene har 20 ganger så høy hyppighet av navnet som det andre kjønn (dvs. man har maks 5 prosent sjanse til å tilordne feil kjønn). For navn med bare tre bokstaver kreves det at det kun er dette navnet som oppgis. Navn med færre enn tre bokstaver benyttes ikke. I tillegg benyttes det et søk etter særegne navnedeler som kan bestemme kjønn.

- C. For de enhetene som har **kjonn** lik blank og **fornavn** ulik blank, letes det i **fornavn** etter tekststrenger som kan bestemme kjønn.
- D. For enheter med **kjonn** lik 'Q' settes **status** lik 'M'. For enheter med **kjonn** lik 'P' og **status** lik 'K' settes **status** lik 'L'.

---

8 Ved denne utvelgelsen risikerer man å klassifisere enkelte personer med ugyldig fødselsnummer som har oppgitt **bokom** eller **postnr** som bosatt i utlandet, dvs. de tas ikke med i statistikken over registrerte ofre som bor i Norge. Dette i motsetning til de enhetene med både uoppgitt **bokom**, **postst** og **postnr** og ugyldig fødselsnummer som vil stå som "uoppgitt" i de fleste sammenhenger. Imidlertid vil treff på søk etter navn på land og uoppgitt **bokom** eller **postnr** indikere at de oppgitte verdiene i **bokom** eller **postnr** er usikre. Hvis **postnr** er uoppgitt, mens **bokom** er oppgitt (med ugyldig fnr), kommer **bokom** fra et funn av en tekststreng i **postst** som er et norsk kommunenavn. Siden landnavn er mer uvanlige bokstavkombinasjoner enn enkelte kommunenavn, antar man her at det er sikrere å kode etter landnavn. Hvis **postnr** er oppgitt, men ikke **bokom**, betyr det at **postnr** ikke har gitt noen gyldig kommunenummer i **bokom**, og det er da grunn til å stille spørsmålsteget ved gyldigheten av **postnr**.

For registrerte ofre som er under myndighetsalder, skal det – ifølge politiets instruksjer – alltid registreres en verge eller foresatt som offer i tillegg til den mindreårige. De foresatte/vergene skal ikke regnes med i statistikken over ofre.

- E. Hvis records med **alder** > 17, **kjonn** lik '1', '2' eller blank og **status** = 'K' har makkerer med hensyn til **polkam**, **saksnr** og **regaar**, der makkeren har **kjonn** lik '1', '2' eller blank, **status** = 'K' og enten  $0 \leq \text{alder} < 18$  eller **alder** lik blank og **lovbrudd** lik '1402', '1403', '1404' eller '1418', settes **status** = 'V'. Hvis det er flere makkerer som kan få **status** = 'V', velges så mange makkerer som det er makkerer med **alder** < 18. Er det flere med **alder** > 17 enn **alder** < 18 velges de med høyest **alder**.<sup>9</sup> Hvis det er færre med **alder** > 17 enn med **alder** < 18, men det i tillegg er noen makkerer med **alder** lik blank, settes **status** lik 'V' for så mange tilfeldig valgte makkerer med **alder** lik blank at antallet med **status** lik 'V' blir likt antallet med **alder** < 18.
- F. Hvis **kjonn** lik 'P', 'Q' eller 'F' settes **alder**, **fdato** og **pnr** lik blank.

## 07 Merking av enheter

SAS-program: \$KRIMINAL/offer/prog/bearbeid/p070.sas

Inndatasett: \$KRIMINAL/offer/arbeid/v060.ssd04

Innfil: \$KRIMINAL/felleskat/lovbrudd.txt  
\$FELLES/sasprog/fnrsjekk.sas

Utdatasett: \$KRIMINAL/offer/data/g[**aar**].ssd04

Loggfil: \$KRIMINAL/offer/logger/logg070.txt

Saksbehandler må oppgi først i programmet hvilken årgang [**aar**] som skal legges ut på data-katalogen.

Hver person som er offer (dvs. **status** lik 'K') skal merkes i det groveste lovbruddet de er offer for. Det beregnes også hvor mange ganger hver person har vært offer

- A. Blant enheter med gyldig fødselsnummer og makkerer med hensyn til **fdato** og **pnr**, settes **merkoff** lik '1' for den enheten som har det groveste lovbruddet (se katalog over lovbrudd, vedlegg 8.2). For enheter med ugyldig fødselsnummer og **kjonn** lik '1', '2' eller blank settes **merke** lik '1'.
- B. Hver enhet med **merkoff** lik '1' og gyldig fødselsnummer får påført **lovant**, som settes lik antall makkerer med hensyn til **fdato** og **pnr** + 1. For enheter med **merkoff** lik '1' og ugyldig fødselsnummer settes **lovant** lik 1.

Hvert lovbrudd merkes, og det beregnes hvor mange ofre det er i hvert lovbrudd

- C. Blant enheter med samme verdi for **polkam**, **aar**, **saksnr** og **lovbrudd** og **status** lik 'K', skal én tilfeldig enhet merkes ved at **merklov** settes lik '1'.

---

<sup>9</sup> Med denne utvelgelsen er man ikke sikker på å velge de riktige vergene blant de personene med **alder** > 17, men man kan anta at der det er flere ofre og verger i samme lovbrudd, vil ofrene som ikke har verge ha **alder** i nærheten av de ofrene som har verge (under 18 år), mens vergene vil ha **alder** som er nærmere en generasjon høyere enn ofrene. Antall aktuelle tilfeller der feil person kan velges som verge er et ensifret tall for alle produserte årganger.

D. For de enheter som har **merklov** lik '1' settes **offant** lik antall personer som var offer for lovbruddet. Det er antall dubletter med hensyn til **polkam**, **aar**, **saksnr** og **lovbrudd**, med **kjonn** lik '1', '2' eller "", og med **status** lik 'K'

**Tabell 8. Filbeskrivelse av det ferdig bearbejdede offermaterialet, g[aar].ssd04**

Kortnavn	Beskrivelse	Format
<b>polkam</b>	Politikammer	char2
<b>aar</b>	Registreringsår	num2
<b>saksnr</b>	Saksnummer	char6
<b>delsak</b>	Delsaksbokstav	char2
<b>regdato</b>	Registreringsdato, maskinelt tilordnet og 100 prosent oppgitt	num4
<b>statbok</b>	Statistikkbokstav ('K' = forbrytelse, 'F' = forseelse)	char1
<b>type</b>	Lovbruddstype, egendefinert ('1' = forbrytelse, '0' = forseelse)	char1
<b>lovbrudd</b>	Lovbruddskode fra STRASAK	char4
<b>modus</b>	Modus	char4
<b>sone</b>	Gjerningsstedssone	char4
<b>gjkom</b>	Gjerningskommune	char4
<b>gjfylke</b>	Gjerningsfylke, '01'-'23' + blank	char2
<b>gjaar</b>	Gjerningsår	num4
<b>gjmd</b>	Gjerningsmåned	num2
<b>gjdag</b>	Gjerningsdag i måneden	num2
<b>postnr</b>	Postnummer	char4
<b>postst</b>	Poststed	char15
<b>status</b>	Status ('K' = offer, 'M' = melder, 'V' = verge, 'L' = politi som offer)	char1
<b>fdato</b>	Fødselsdato	ddmmyy
<b>pnr</b>	Personnummer	char5
<b>slnavn</b>	Slektsnavn	char50
<b>fornavn</b>	Fornavn	char50
<b>boland</b>	Hvilket land bostedsadressen oppgir, Statistisk sentralbyrås landkode	char3
<b>bokom</b>	Bokommune 1. januar, eventuelt 31. desember, Statistisk sentralbyrå sin standard for	char4
<b>stat</b>	Statsborgerskap 1. januar, evt 31. desember, i statistikkåret	char3
<b>botid</b>	Botid i hele år	num2
<b>bydel</b>	Bydel i Oslo	char2
<b>invkat</b>	Innvandrerkategori	char1
<b>ib</b>	Innvandrerstatus ('1' = med i innvanderbefolkningen, '0' = ikke med, blank = ubestemt)	char1
<b>ibland</b>	Bakgrunnsland, Statistisk sentralbyrå sin standard for landkode	char3
<b>ibvereg</b>	Verdensregion, bakgrunnsland	char1
<b>alder</b>	Alder på gjerningstidspunkt	num3
<b>kjonn</b>	Kjønn ('1' = mann, '2' = kvinne, 'P' og 'Q' = politi, 'F' = foretak, blank = ukjent)	char1
<b>merklov</b>	Merking av hvert lovbrudd	char1
<b>merkoff</b>	Merking av offer	char1
<b>lovant</b>	Antall lovbrudd hver person har vært offer for	num3
<b>offant</b>	Antall ofre som er registrert i hvert lovbrudd	num3

### 3.5 Dokumentasjon av bearbejdingen

Under kjøringen av hvert SAS-program skrives det ut en del opplysninger til en logg. Opplysningene kan være antall omkodinger, antall slettinger, antall enheter med visse variabelverdier, antall enheter der en logisk test slo til eller ikke og så videre. Loggene legges i katalogene '\$KRIMINAL/anmeldt/logger' og '\$KRIMINAL/offer/logger' med navn som korresponderer med navnet på SAS-programmet som opprettet dem (se filstrukturen i vedlegg 7.1).

Loggene skal samles til ett tekstdokument ved bearbejdingens slutt. Dette samledokumentet lagres permanent i arkiv-katalogen, med et navn som forteller hvilken årgang loggene refererer til. For å

unngå forvirring med hensyn til bearbeidingens status bør enkeltloggene slettes før bearbeidingen av ny årgang starter.

En del av SAS-programmene inneholder også generering av enkle tabeller for de variablene som behandles i programmet. Typisk lages det en frekvenstabell over variabelverdier for en variabel før maskinell omkoding, og en tilsvarende frekvenstabell etter omkoding. Tabellene gir en mulighet for kvalitetssikring umiddelbart etter kjøring av et program. Dette gjelder først og fremst den logiske bearbeidingen av datamaterialet, men også tekniske problemer vil kunne bli avdekket ved å sammenligne tabellene. Disse tabellene danner en omfattende dokumentasjon over bearbeidingen, men blir kun lagt i SAS sin OUTPUT og forsvinner når SAS avsluttes. Alle tabellene må derfor skrives ut på papir, eventuelt lagres manuelt, hvis det er ønskelig å ta vare på resultatene for ettertiden.

## 4. Tabellproduksjon

### 4.1 Generelt

Tabeller over anmeldte lovbrudd vil i stor grad benytte seg av avledede variable eller kategorisering av variabelverdier. Tidligere ble omkodede verdier lagt på datasettet. Siden kategorisering og omkoding endrer seg over tid, ikke minst som en følge av at politiet videreutvikler sitt STRASAK-system, er det mest hensiktsmessig å foreta disse beregningene på nytt hver gang de aktuelle tabellene lages. Omkoding og kategorisering benytter derfor SAS-formater i utstrakt grad. Det er imidlertid de underliggende *katalogene* som må være oppdatert (se vedlegg 8.2), mens formatene opprettes på nytt ved hver tabellkjøring. Programkoden til oppretting av SAS-formatene kan dermed gjenbrukes (kopieres fra tidligere tabeller) uten at dette forhindrer oppdatering av omkodingen/kategoriseringen.

Statistikken inneholder sensitive data og behandles i henhold til *Sikkerhetshåndboka* (Statistisk sentralbyrå 1998a). For publisering og spredning av informasjon der personer er enhet (gjelder bare offerstatistikk), gjelder spesielt kapitlene 6 og 7. Det er imidlertid ikke alle tabeller som har personer som enhet. Når lovbrudd er enhet kan publisering skje uten at krav om anonymitet for personer er aktuelt.

Vedlegg 7.1 viser plassering av de mest sentrale tabellprogrammene under filstrukturen.

### 4.2 Utvelgelse

Når det lages tabeller fra det endelige SAS-datasettet skjer det en utvelgelse av dataene som er avhengig av hvilken enhet tabellen viser. Denne utvelgelsen skal ligge i hvert enkelt tabellprogram.

**Tabell 9. Utvelgelse ved bruk av forskjellige enheter i tabellene**

Enhet	Utvalg	Merknad
<i>For anmeldte lovbrudd:</i>		
Lovbrudd	<b>lovbrudd</b> < '9200', <b>statbok</b> = 'K', 'F', <b>polkam</b> ulik '56' og <b>saksbok</b> lik blank	Denne utvelgelsen er gjort for de siste årgangene i bearbeidingen
<i>For registrerte ofre:</i>		
Offerhendelser	<b>status</b> = 'K'	
Offerhendelser, personer	<b>status</b> = 'K' og <b>kjonn</b> lik '1', '2' eller blank	Vanligste enheten
Offerhendelser i forhold til folketall	<b>status</b> = 'K' og gyldig fødselsnummer	
Personer registrert som offer	<b>status</b> = 'K' og <b>merkoff</b> = '1'	Problematisk med mange uoppgett f-numre
Anmeldere	<b>status</b> ulik 'V' og <b>merklov</b> = '1'	

### 4.3 Begrensninger i bruk av variable

Ved tabellproduksjon vil en del variable ikke kunne benyttes ved alle årganger, enten fordi de ikke er tatt med, eller fordi kvaliteten er så dårlig at det ikke er faglig forsvarlig. Variablene er gyldige hvis ikke annet er oppgitt i tabell 10.

I tillegg er det noen variable som har dårlig kvalitet, men som kan benyttes på et visst summarisk nivå. Kombinasjonmulighetene og aggregeringsnivåene er imidlertid for mange til å gi en enkel oversikt over denne problemstillingen. Dette må saksbehandler ta stilling til ved det enkelte tabelldesign og den enkelte tabellkjøringen.

**Tabell 10. Variable med begrenset bruksområde**

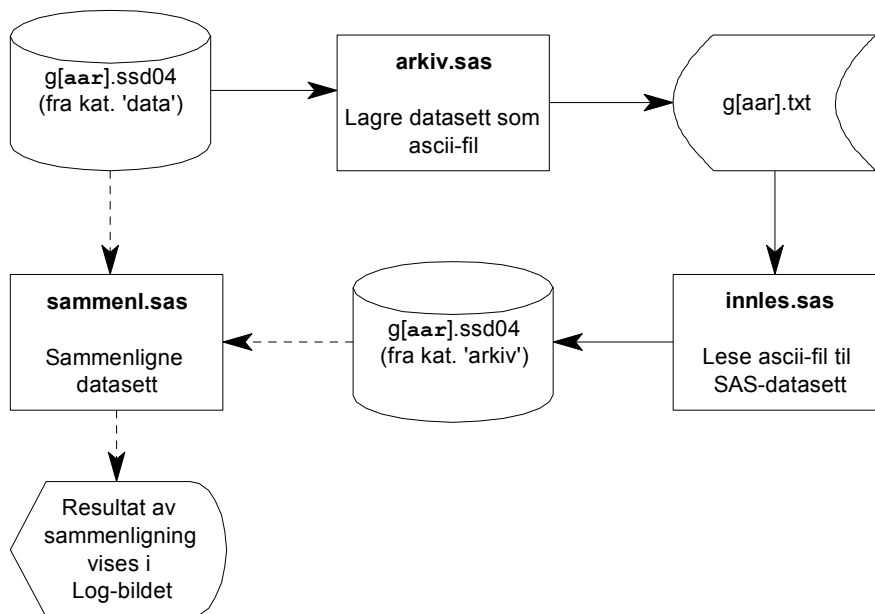
Variabel	Periode	Manglende gyldighet	Merknad
<i>Anmeldte lovbrudd</i>			
<b>modus</b>			Bare 25 prosent har oppgitt modus
<b>polkam</b>	1991–1995	Mangler bare Sysselmannen	Sysselmannen på Svalbard var ikke koblet til STRASAK før 1995
<b>lovbrudd</b>	1991–1994	Samme kode endret tekst (utuktig omgang med mindreårig)	<b>Lovbrudd</b> = '1403' må kodes om til '1499' hvis <b>regaar</b> < 1995
<i>Registrerte ofre</i>			
<b>polkam</b>	1993–1995	Mangler bare Sysselmannen	Sysselmannen på Svalbard var ikke koblet til STRASAK før 1995
<b>lovbrudd</b>	1993–1994	Samme kode endret tekst (utuktig omgang med mindreårig)	<b>Lovbrudd</b> = '1403' må kodes om til '1499' hvis <b>regaar</b> < 1995
<b>bokom</b>	1993–1994/1995	For lav andel med gyldig fødselsnr	Fnr mindre oppgitt for tidligere årganger
<b>stat</b>	1993–1994/1995	For lav andel med gyldig fødselsnr	Fnr mindre oppgitt for tidligere årganger
<b>merkoff</b>	1993–1994/1995	For lav andel med gyldig fødselsnr	Fnr mindre oppgitt for tidligere årganger
<b>merklov</b>	1993–	Lav andel med fnr og uhensiktsmessig registreringsrutiner	Materialet er lite egnet til å beskrive problemstillinger knyttet til lovbrudd med flere ofre. Enkeltsaker registreres
<b>offant</b>	1993–	Lav andel med fnr og uhensiktsmessig registreringsrutiner	Materialet er lite egnet til å beskrive problemstillinger knyttet til lovbrudd med flere ofre. Enkeltsaker registreres
<b>lovant</b>	1993–1994/1995	For lav andel med gyldig fødselsnr	Fnr mindre oppgitt for tidligere årganger
<b>modus</b>	1993–		Bare 25 prosent har oppgitt modus
<b>botid</b>	1993–1994	Finnes ikke	
<b>botid</b>	1995–		Lik <b>alder</b> for de som er født i Norge
<b>bydel</b>	1993–1994	Finnes ikke	
<b>bydel</b>	1995–		Bare for Oslo
<b>invkat</b>	1993–1994	Finnes ikke	
<b>ib</b>	1993–1994	Finnes ikke	
<b>ibland</b>	1993–1994	Finnes ikke	
<b>ibvereg</b>	1993–1994	Finnes ikke	

## 5. Permanent lagring av data

### 5.1 Lagring av ferdig revidert materiale

For å gardere seg mot eventuelle nye formater eller tekniske verktøy skal alle vitale data lagres som rene ascii-filer. Dette gjelder spesielt de ferdige bearbejdede årgangsfilene, men omfatter også rådata og bearbejdingsdokumentasjon.

Alle årganger på de to statistikkområdene følger samme struktur på de flate filene. Alle variable vil dermed ha sin plass på filen, men for de årganger som ikke omfatter variabelen vil plassen være tom. Tabell 10 kan brukes for å se hvilke variable som er gyldig for den enkelte årgang. Fordelen med å opprette strukturelt like ascii-filer for alle årganger er at det kan benyttes ett felles SAS-program for innlesing av flat fil til SAS-datasett og ett SAS-program for lagring av ascii-filene for alle årganger for henholdsvis anmeldte lovbrudd og registrerte ofre.



Figur 8. Logisk dataflyt, arkivering og innlesing av årgangsdata

Utfilene skal være i overensstemmelse med beskrivelsen i tabell 8 og 9. Ikke alle årganger inneholder alle variable, men da settes tilsvarende posisjoner som blanke. Ved bruk av arkivdata vil derfor en del variable ha bare blanke verdier.

## For anmeldte lovbrudd

SAS-program: \$KRIMINAL/anmeldt/prog/arkiv/arkiv.sas  
 Inndatasett: \$KRIMINAL/anmeldt/data/g[**aar**].ssd04  
 Utfil: \$KRIMINAL/anmeldt/arkiv/a[**aar**].txt

**Tabell 11. Filbeskrivelse av årgangfilene ved lagring av flat fil – anmeldte lovbrudd**

Posisjon	Beskrivelse	Kortnavn ved innlesning
1–2	Politidistrikt	polkam
3–4	År	aar
5–10	Saksnummer	saksnr
11–12	Saksbokstav	saksbok
13–16	Sekvens	sekvens
17–24	Registreringsdato	regdato
25	Statistikkbokstav	statbok
26–29	Lovbrudd	lovbrudd
30–33	Modus	modus
34–37	Sone	sone
38–41	Gjerningskommune	gjkom
42–49	Gjerningsdato start	gjdato1
42–45	Gjerningsår start	gjaar1
46–47	Gjerningsmåned start	gjmd1
48–49	Gjerningsdag start	gjdag1
50–57	Gjerningsdato slutt	gjdato2
50–53	Gjerningsår slutt	gjaar2
54–55	Gjerningsmåned slutt	gjmd2
56–57	Gjerningsdag slutt	gjdag2
58	Tidel	tidel



**For registrerte ofre**

SAS-program: \$KRIMINAL/offer/prog/arkiv/arkiv.sas

Inndatsett: \$KRIMINAL/offer/data/g[**aar**].ssd04Utfil: \$KRIMINAL/offer/arkiv/a[**aar**].txt**Tabell 12. Filbeskrivelse av årgangfilene ved lagring av flat fil – registrerte ofre**

<b>Posisjon</b>	<b>Beskrivelse</b>	<b>Kortnavn ved innlesing</b>
1–2	Politikammer	<b>polkam</b>
3–4	Registreringsår	<b>aar</b>
5–10	Saksnummer	<b>saksnr</b>
11–12	Delsaksbokstav	<b>delsak</b>
13–20	Registreringsdato, maskinelt tilordnet og 100 prosent oppgitt	<b>regdato</b>
13–16	Registreringsår	<b>regaar</b>
17–18	Registreringsmåned	<b>regmnd</b>
19–20	Registreringsdag	<b>regdag</b>
21	Statistikkbokstav ('K' = forbrytelse, 'F' = forseelse)	<b>statbok</b>
22–25	Lovbruddskode fra STRASAK	<b>lovbrudd</b>
26–29	Modus	<b>modus</b>
30–33	Gjerningsstedssone	<b>sone</b>
34–37	Gjerningskommune	<b>gjkom</b>
38–39	Gjerningsfylke, '01'–'23' + blank	<b>gjfylke</b>
40–43	Gjerningsår	<b>gjaar</b>
44–45	Gjerningsmåned	<b>gjmd</b>
46–47	Gjerningsdag i måneden	<b>gjdag</b>
48–51	Postnummer	<b>postnr</b>
52–66	Poststed	<b>postst</b>
67	Status ('K' = offer, 'M' = melder, 'V' = verge, 'L' = politi som offer)	<b>status</b>
68–75	Fødselsdato	<b>fdato</b>
68–71	Fødselsår	<b>faar</b>
72–73	Fødselsmåned	<b>fmnd</b>
74–75	Fødselsdag	<b>fdag</b>
76–80	Personnummer	<b>pnr</b>
81–130	Slektsnavn	<b>slnavn</b>
131–180	Fornavn	<b>fornavn</b>
181–183	Hvilket land bostedsadressen oppgir, Statistisk sentralbyrås landkode	<b>boland</b>
184–186	Bokommune 1. januar, Statistisk sentralbyrå sin standard for kommunenumre	<b>bokom</b>
187–189	Statsborgerskap 1. januar	<b>stat</b>
190–191	Botid i hele år	<b>botid</b>
192–193	Bydel i Oslo	<b>bydel</b>
194	Innvandrerkategori	<b>invkat</b>
195	Innvandrerstatus ('1' = med i innvandrerbef., '0' = ikke med, blank = ubestemt)	<b>ib</b>
196–198	Bakgrunnsland, Statistisk sentralbyrå sin standard for landkode	<b>ibland</b>
199	Verdensregion, bakgrunnsland	<b>ibvereg</b>
200–202	Alder på gjerningstidspunkt	<b>alder</b>
203	Kjønn ('1' = mann, '2' = kvinne, 'P' og 'Q' = politi, 'F' = foretak, blank = ukjent)	<b>kjonn</b>
204	Merking av hvert lovbrudd	<b>merklov</b>
205	Merking av offer	<b>merkoff</b>
206–208	Antall lovbrudd hver person har vært offer for	<b>lovant</b>
209–211	Antall ofre som er registrert i hvert lovbrudd	<b>offant</b>
212	Lovbruddstype, egendefinert ('1' = forbrytelse, '0' = forseelse)	<b>type</b>

## 5.2 Lagring av andre filer permanent

I tillegg til de ferdig reviderte dataene i hver statistikkårgang, skal en del andre ascii-filer lagres permanent. Dette gjøres delvis for å kunne rekonstruere revisjonen, og delvis for å kunne dokumentere denne revisjonen.

Alle rådatafiler fra Politiets datatjeneste lagres permanent. Disse filene er utgangspunktet for statistikkene og lar seg ikke rekonstruere hos Politiets datatjeneste ved et senere tidspunkt.<sup>10</sup> Filene fra Politiets datatjeneste lagres under katalogen \$KRIMINAL/anmeldt/arkiv og gis navn etter formen PD[**aar**].txt

Hvert SAS-program under revisjonsrutinene skriver ut en loggfil. Disse loggfilene samles i ett dokument når én statistikkårgang er ferdig revidert. Denne loggen lagres permanent for senere dokumentasjon under katalogen \$KRIMINAL/anmeldt/arkiv og \$KRIMINAL/offer/arkiv og gis navn etter formen logg[**aar**].txt.

## 5.3 Innlesing av ascii-filer til SAS-datasett

Det er bare ascii-filene \$KRIMINAL/anmeldt/arkiv/g[**aar**].txt og \$KRIMINAL/offer/arkiv/g[**aar**].txt som inneholder de ferdig reviderte dataene det er aktuelt å lese inn på nytt via et eget SAS-program. Samme program under hvert statistikkområde kan benyttes for innlesing av alle årgangers ascii-filer til SAS-datasett. I tillegg må det spesifiseres hvilken årgang som det leses inn fra øverst i programmet. Hvis det ønskes eksakt like datasett som det opprinnelige, må de variable som ikke skal med spesifiseres i SAS-programmet (ved hjelp av drop-statement).

### For anmeldte lovbrudd

SAS-program: \$KRIMINAL/anmeldt/prog/arkiv/innles.sas  
Innfil: \$KRIMINAL/anmeldt/arkiv/g[**aar**].txt  
Utdatasett: \$KRIMINAL/anmeldt/arkiv/g[**aar**].ssd04

Utdatasettet legges under arkiv-katalogen, og må manuelt flyttes til datakatalogen for eventuelt videre bruk. Dette er gjort som en ekstra sikkerhet mot uønsket sletting av datasett, og der er da også mulig å sammenligne de opprinnelige datasettene med arkivmateriale når dette opprettes. Formaterte datasett under arkiv-katalogen skal normalt ikke forekomme, og bør slettes umiddelbart etter at sammenligninger er foretatt, og eventuelt dataene er kopiert over i data-katalogen

Det er laget et SAS-program som sammenligner arkivmaterialet med opprinnelig datasett. Dette kan generelt benyttes til å sammenligne datasett, se figur 8. Kopi av programmet er lagt på området for både anmeldte lovbrudd og registrerte ofre. Det skrives ut en rapport i *Output* i SAS som gir resultatene av sammenligningen.

---

<sup>10</sup> Politiets datatjeneste foretar en sanering av saker i STRASAK minst én gang i året. Sanering av saker følger flere kriterier, blant annet alder på saken og grovhet på lovbrudd. I prinsippet tas alt materiale kopi av. I praksis vil det imidlertid være et svært omfattende arbeid å skulle rekonstruere det datamaterialet som ligger til grunn for én årgang av statistikken over ferdig etterforskede lovbrudd.

SAS-program: \$KRIMINAL/anmeldt/prog/arkiv/sammenl.sas  
Inndatasett: \$KRIMINAL/anmeldt/arkiv/g[**aar**].ssd04  
\$KRIMINAL/anmeldt/data/g[**aar**].ssd04

### For registrerte ofre

SAS-program: \$KRIMINAL/offer/prog/arkiv/innles.sas  
Innfil: \$KRIMINAL/offer/arkiv/g[**aar**].txt  
Utdatasett: \$KRIMINAL/offer/arkiv/g[**aar**].ssd04

Utdatasettet legges under arkiv-katalogen, og må manuelt flyttes til datakatalogen for eventuelt videre bruk. Dette er gjort som en ekstra sikkerhet mot uønsket sletting av datasett, og der er da også mulig å sammenligne de opprinnelige datasettene med arkivmateriale når dette opprettes. Formaterte datasett under arkiv-katalogen skal normalt ikke forekomme, og bør slettes umiddelbart etter at sammenligninger er foretatt, og eventuelt dataene er kopiert over i data-katalogen

Det er laget et SAS-program som sammenligner arkivmaterialet med opprinnelig datasett. Dette kan generelt benyttes til å sammenligne datasett, se figur 8. Kopi av programmet er lagt på området for både anmeldte lovbrudd og registrerte ofre.

SAS-program: \$KRIMINAL/offer/prog/arkiv/sammenl.sas  
Inndatasett: \$KRIMINAL/offer/arkiv/g[**aar**].ssd04  
\$KRIMINAL/offer/data/g[**aar**].ssd04

Det skrives ut en rapport i *Output* i SAS som gir resultatene av sammenligningen.

## 6. Referanser

Alle referanser finnes i Statistisk sentralbyrås bibliotek. Unntaket er rundskriv fra Justisdepartementet som finnes i "STRASAK i et nøtteskall" som oppbevares ved fagseksjon.

Berby, H. (1999) "Datalagring på UNIX i Statistisk sentralbyrå: Retningslinjer", *Interne dokumenter*, nr. 1, Statistisk sentralbyrå. 3. reviderte utgave.

Bjørsvik, A. og H.-K. Torvbråten (1999) "Brukerbeskrivelse for Dokumentasjonssystem i Statistisk sentralbyrå, DataDok V1.1", *Interne dokumenter*, nr. 6, Statistisk sentralbyrå.

Gundersen, F. (1999) "Statistikk over etterforskede lovbrudd. Dokumentasjon", *Notater*, nr. 52, Statistisk sentralbyrå.

Justisdepartementet (1994) *Rundskriv G-73/94*

Løvkvist, H. (1997) "Feilidentifisering og imputering i forbindelse med revisjon av administrative registre", *Interne dokumenter*, nr. 22, Statistisk sentralbyrå.

Olaussen, L. P. (1996) "Kriminalstatistikk og virkelighet", *PHS Småskrifter 1996:1*, Oslo: Politihøgskolen.

Sentralkontoret for folkeregistrering (1996) *Kodeliste for land/statsborgerskap*, 1. februar.

Statistisk sentralbyrå (1994) *Standard for kommuneklassifisering 1994*, NOS C 192, Statistisk sentralbyrå.

Statistisk sentralbyrå (1995) *Håndbok i prosjektstyring*, Statistisk sentralbyrås håndbøker, nr. 58, Statistisk sentralbyrå.

Statistisk sentralbyrå (1997) *Håndbok i Edb-metode. Edb-metode 96: Minimetode for vedlikehold og utvikling av Edb-systemer i Statistisk sentralbyrå*, Statistisk sentralbyrås håndbøker, nr. 63, Statistisk sentralbyrå.

Statistisk sentralbyrå (1998a) *Håndbok i datasikkerhet og fysisk sikring*, Statistisk sentralbyrås håndbøker, nr. 45, revidert utgave, november 1998, Statistisk sentralbyrå.

Statistisk sentralbyrå (1998b) *Kriminalstatistikk 1996*, NOS C 466, Statistisk sentralbyrå.

Statistisk sentralbyrå (1998c) *Håndbok i utvikling av statistikkssystemer. Med vekt på IT-metode*, Statistisk sentralbyrås håndbøker, nr. 65, Statistisk sentralbyrå.

Statistisk sentralbyrå (1998d) *Håndbok i datarevisjon*, Statistisk sentralbyrås håndbøker, nr. 66, Statistisk sentralbyrå.

Statistisk sentralbyrå (1999) *Regionale inndelinger: En oversikt over standarder i norsk offisiell statistikk*, NOS C 513, Statistisk sentralbyrå.

## 7. Vedlegg

### 7.1 Filstruktur

I kriminalstatistikken er det hovedsakelig benyttet en filstruktur slik det er anbefalt i Berby (1999). All dokumentasjon ligger på fellesdisken på lokalnettet, mens alle SAS-datasett, kataloger og programmer ligger på UNIX.

#### På fellesdisk på lokalnettet

Dokumentasjonen for statistikk over anmeldte lovbrudd har et område på en fellesdisk med adresse Q:\dok\kriminal\anmeldt. Statistikk over registrerte ofre har ikke eget område, men inngår i dokumentasjonen over anmeldte lovbrudd. Alle dokumenter som inngår i dette notatet er samlet i en underkatalog som heter "notat".

#### Filstruktur på lokalnettet:

Stamme	Nivå 1	Nivå 2	Nivå 3	Beskrivelse
Q:\DOK\Kriminal	anmeldt	notat	tekst.doc	Dette dokumentet, Microsoft Word-format Figur 1, saksgang for lovbrudd, er i EPS-format, og kan redigeres i Adobe Illustrator. Figurene 2 til 5 i notatet, dataflyt og aldersberegning, er i formatet til Micrografx ABC FlowCharter
			figur1.eps	
			figur2.af3	
			revisjon	pol_adr.doc pol_lab.doc foelge.doc
		diverse		Diverse dokumenter som saksbehandler finner det formålstjenlig å lagre under dokumentasjonskatalogen
	eforsk			En annen kriminalstatistikk
	reaksjon			En annen kriminalstatistikk
	fengsel			En annen kriminalstatistikk
X:\350\kriminal				Fellesområde for kriminalstatistikken, brukes til faglige prosjekter (NOS, Rapporter o.l.)

#### På UNIX-tjeneren

Kriminalstatistikk har hovedstamme \$KRIMINAL, men det er lagt inn ett nivå mer enn standarden anbefaler. Filstrukturen under \$KRIMINAL vil derfor omfatte mer enn statistikken over anmeldte lovbrudd og registrerte ofre, men bare den delen av strukturen som berører produksjon av disse statistikkene er tatt med her. I tabellen er kataloger skrevet med uthevet skrift, mens fysiske filer er skrevet med normal skrift.

SAS skal alltid startes fra 'prog'-katalogen. Dette for å lese filen autoexec.sas under oppstart. Denne filen inneholder referanser til de ulike katalogene som benyttes i de enkelte programmene.

## Filstruktur på lokalnettet:

Nivå 1	Nivå 2	Nivå 3	Nivå 4	Beskrivelse
anmeldt	prog	autoexec.sas		Inneholder "librefs" til alle kataloger i denne filstrukturen.
		bbasINN.sas		Leser flate befolkningsfiler inn i SAS-datasett
		<b>bearbeid</b>	p010.sas – p110.sas	SAS-programmer som utfører selve bearbeidingen av datamaterialet
		<b>arkiv</b>	arkiv.sas innles.sas sammenl.sas	SAS-prog. som legger ut det ferdige datasettene som flate filer SAS-prog som leser inn flate filer u/tilbakefall til SAS-datasett SAS-prog som sammenligner to datasett
		<b>NOS</b>	tab01.sas – tab04.sas	Alle SAS-programmene som lager tabeller til NOSen
		<b>oppdrag</b>	A_*.sas E_*.sas O_*.sas	Tabellprogrammer for eksterne oppdrag, både engangsoppdrag og gjentatte (f.eks. "Hjulet", Interpol). Programnavnet starter med en bokstav som forteller hvilket statistikkområdet det henter data fra.
	<b>data</b>	g1990.ssd04 – g[ <b>aar</b> ].ssd04		De ferdig bearbeidede årgangsfilerne som danner grunnlag for tabeller. Lagres permanent.
	<b>inndata</b>	PD[ <b>aar</b> ].txt		Rådatafiler fra Politiets datatjeneste
	<b>arbeid</b>	v010.ssd04 – v030.ssd04		Datasett som lagres under hele bearbeidingen, men som kan slettes før ny årgang kommer inn.
	<b>logger</b>	logg010.txt – logg040.txt		Logger som lages ved bearbeidingen. Samles i ett dok. som lagres i arkiv-katalogen. Enkeltloggene slettes før ny årgang bearbeides.
	<b>arkiv</b>	g[ <b>aar</b> ].txt PD[ <b>aar</b> ].txt logg[ <b>aar</b> ].txt		Ascii-filer av ferdig reviderte datasett Rådataene for de enkelte årgangene fra PD Hele loggen over bearbeidingen i statistikkåret
	offer	prog	autoexec.sas	
bbasINN.sas				Leser flate befolkningsfiler inn i SAS-datasett
Soundex.sas				makro for koding av tekststrenger til "lydskrift"
<b>bearbeid</b>			p010.sas – p070.sas	SAS-programmer som utfører selve bearbeidingen av datamaterialet
		<b>arkiv</b>	arkiv.sas innles.sas sammenl.sas	SAS-prog. som legger ut det ferdige datasettene som flate filer SAS-prog som leser inn flate filer u/tilbakefall til SAS-datasett SAS-prog som sammenligner to datasett
		<b>NOS</b>		SAS-programmer som lager tabeller til NOSen – i dag tomt
		<b>oppdrag</b>	A_*.sas E_*.sas O_*.sas	Tabellprogrammer for eksterne oppdrag, både engangsoppdrag og gjentatte (f.eks. "Hjulet", Interpol). Programnavnet starter med en bokstav som forteller hvilket statistikkområdet det henter data fra.
<b>data</b>		g1993.ssd04 – g[ <b>aar</b> ].ssd04		De ferdig bearbeidede årgangsfilerne som danner grunnlag for tabeller. Lagres permanent.
<b>inndata</b>		BBAS[ <b>aar</b> ].txt firma.ssd04 navn.txt kortnavn.txt		Data som benyttes som input i produksjonen
<b>arbeid</b>		v010.ssd04 – v060.ssd04		Datasett som lagres under hele bearbeidingen, men som kan slettes før ny årgang kommer inn.
<b>logger</b>		logg010.txt – logg070.txt		Logger som lages ved bearbeidingen. Samles i ett dok. som lagres i arkiv-katalogen. Enkeltloggene slettes før ny årgang bearbeides.
<b>arkiv</b>		g[ <b>aar</b> ].txt PD[ <b>aar</b> ].txt logg[ <b>aar</b> ].txt		Ascii-filer av ferdig reviderte datasett Rådataene for de enkelte årgangene fra PD Hele loggen over bearbeidingen i statistikkåret
felleskat	<b>kat</b>		Kataloger som er spesifikke for statistikk over registrerte ofre	
	lovbrudd.txt sone.txt polkam.txt		Kataloger, eventuelt datasett, som er felles for flere statistikkområder under kriminalstatistikk. Ikke alle er vist her, se vedlegg 7.2.	
fodeland	g[ <b>aar</b> ].txt g[ <b>aar</b> ].ssd04		Rådata fra seksjon 320 ("fødelandsfilen") SAS-datasett tilsvarende fødelandsfilen	
reaksjon			En annen statistikk under \$KRIMINAL	
fengsel			En annen statistikk under \$KRIMINAL	
eforsk			En annen statistikk under \$KRIMINAL	

## 7.2 Kataloger

Programmer og formatdeklarasjoner skal ikke oppdateres med mindre det skal foretas kvalitative endringer i bearbeidingsrutinene. Katalogene derimot representerer den delen av bearbeidingsrutinene som fortløpende skal oppdateres. Av disse er det spesielt lovbruddskatalogen som hyppig blir oppdatert med nye lovbruddskoder fra STRASAK-rådet.

### Lovbruddskatalog

Fysisk fil: \$KRIMINAL/felleskat/lovbrudd.txt

Pos	Forklaring	Verdier
1-4	Lovbrudd, statistikkgruppe	'0100' – '9200'
6-6	Lovbruddsgruppe	'0' – '9' (den etablerte standarden "Tidelingen")
7-7	Lovbruddskategori	'1' = forbrytelse, '0' = forseelse og 'B' = ikke entydig
9-11	Grovhetsindikator	'010' – '210', høyere indikator er grovere lovbrudd
13-18	Maks strafferamme:	
13-14	Antall år	0-21
16	Antall måneder	0-9
18	Bot er maksimumsstraff	'1' = bot er maksimumsstraff
20-25	"Normal" strafferamme	Antall år 20-21, antall måneder 23, bot som maks 25 (=1), disse opplysningene benyttes foreløpig ikke
27-33	Minimum strafferamme:	
27	Antall år	0-9
29	Antall måneder	0-9
31	Fengsel som minimum	'1' = fengsel er minimumsstraff
33	Bot som minimum	'1' = bot er minimumsstraff
35	Bot i tillegg til fengsel	'1' = bot kan gis som tilleggsstraff til fengsel
37	Påtaleplikt hos politiet	'J' = Lovbruddet er underlagt offentlig påtale 'N' = Lovbruddet er ikke underlagt offentlig påtale
39-42	Ny kode	Hvis lovbruddskoden ikke er gyldig hos politiet kan den erstattes av oppgitte kode (benyttes foreløpig ikke i bearbeidningen)
44-98	Lovbruddstekst	Beskrivelse av lovbruddet, samme tekst som i politiets kodebok

### Gamle kommuner

Fysisk fil: \$KRIMINAL/felleskat/kom\_gml.txt

Pos	Forklaring
1-4	Utgåtte kommunekoder, basert på Statistisk sentralbyrå (1994)
6-9	Gyldige kommunekoder

### Katalog over landnavn

Fysisk fil: \$KRIMINAL/felleskat/landnavn.txt

Pos	Forklaring
1-3	Statistisk sentralbyrå sin landkode
5-45	Navn på alle land i verden + ulike tekststrenger som identifiserer land (f.eks. 'DK-', 'London', 'GERMANY' o.l.)

## Sonekatalog

Fysisk fil: \$KRIMINAL/felleskat/sone.txt

Pos	Forklaring	Verdier
1-2	Politikammer	'01' – '60'
3-6	Sone	
8-9	Fylkesnummer	'01'–'21'
11-14	Kommunennummer	Statistisk sentralbyrås standard kommunenummer
16-21	Bydelsnummer	Fire første sifrene er kommune, to siste er bydel
23-50	Tekst	Beskrivelse av sonen, samme som politiets

## Myndighetskatalog

Fysisk fil: \$KRIMINAL/felleskat/myndigh.txt

Pos	Forklaring	Verdier
1-4	Myndighetskode	Tallkode
6-9	Kommunekoder	Statistisk sentralbyrås standard kommunenummer
11-12	Fylkesnummer	'01'–'21'
14-40	Navn på myndighet	

## Landkatalog

Fysisk fil: \$KRIMINAL/felleskat/land.txt

Pos	Forklaring	Verdier
1-3	Landkode/statsborgerskap fra STRASAK	Bokstavkode
5-7	Statistisk sentralbyrås landkode	Tallkode
9-11	Statistisk sentralbyrås kode for statsborgerskap	Tallkode
13-58	Navn på land	Tekst

## Statsborgerskapskatalog

Fysisk fil: \$KRIMINAL/felleskat/stat.txt

Pos	Forklaring	Verdier
1-3	Statistisk sentralbyrås kode for statsborgerskap	Tallkode
5-33	Navn på statsborgerskap	Tekst
34-74	Navn på land/region	Tekst
75-103	Navn på Verdensdel	Tekst

## Politikammerkatalog

Fysisk fil: \$KRIMINAL/felleskat/polkam.txt

Pos	Forklaring	Verdier
1-2	Politikammer	'01'–'62'
4-7	Kommunennummer	Kommunen som sammenfaller med politidistriktet
9-12	"Fylkesnummer"	'0100'–'2100'
14-32	Navn på politikammer	Tekst



## Katalog over postsoner

Denne skal egentlig befinne seg på \$FELLES/ssbkat/poststed/, men det er ikke etablert faste rutiner for oppdatering, filbeskrivelse og filnavn. Katalogen er derfor kopiert til kriminalstatistikken sitt hjemmeområde, men hvis det opprettes faste rutiner sentralt i Statistisk sentralbyrå, anbefales det å benytte den sentrale versjonen. Kvalitetssikring, oppdatering og dokumentasjon vil dermed ivaretas bedre.

Fysisk fil: \$KRIMINAL/felleskat/post.txt

Pos	Forklaring	Verdier
1-4	Postnummer	'0000'-'9999'
5-39	Navn på poststed	Tekst
40-43	Korresponderende kommunenr	'0100'-'2100'
44-78	Navn på kommune	Tekst

## Kommunekatalog

Fysisk fil: \$FELLES/Statistisk sentralbyråkat/kommune/g1994

Filbeskrivelse: \$FELLES/SSBkat/kommune/kommune.dok

## 7.3 Kontroller i STRASAK

I STRASAK-systemet ligger det inne noen automatiske tester. Denne oversikten er basert på opplysninger fra Politiets datatjeneste 19. september 1995.

### Datotesting

**Regmnd/regaar**, fullstendig **gjdato**, **fdato** og **avgjmnd/avgjaar** er underlagt følgende kriterier:

- Dato skal være numerisk og ikke bestå av bare nuller.
- Måned skal ha verdi 1–12.
- Dag må ha verdi 1–31.
- Dag skal ikke overskride det antall dager som den aktuelle måned har, inkl. test på skuddår.
- Dato skal ikke være større enn dagens dato.

Det tillates å registrere **gjdato** ufullstendig, men med unntak av at

- det ikke er tillatt å registrere bare dag og måned,
- det ikke er tillatt å registrere bare dag og år,
- det ikke er tillatt å registrere bare måned,
- det ikke er tillatt å registrere bare dag.
- månedsdelen kan ikke være større enn 12 og årsdelen ikke større enn regaar

I praksis betyr dette at dato kan oppgis som måned og år eller bare år. Årsdelen vil alltid være oppgitt.

### Lovbruddskoder

- **Lovbrudd** må finnes i tabell i databasen, men her ligger alle kodene som er og har vært i bruk.
- **Lovbrudd** må ikke slutte på «00», som er hovedstatistikkgruppe.
- Noen koder som ikke skal benyttes (de utgåtte) sjekkes ved nyregistrering og nektes benyttet. Dette gjelder imidlertid ikke alle utgåtte koder.
- Statistikkbokstav må være oppgitt
- Statistikkbokstav testes mot **lovbrudd** der noen koder er definert som forseelser og noen som forbrytelser. Manglende oppdateringer forekommer.

### Kommunekoder

Kodene for **gjkom** og **bokom** må finnes i en tabell over kommunekoder, men denne tabellen inneholder både gamle og nye kommunenummer.

Politiets datatjeneste tar forbehold om at feil kan forekomme, til tross for utførte tester. Dette bekreftes også av Statistisk sentralbyrås egne kontroller.

## 7.4 Spesielle variabelverdier

En del variable har tilordnet verdier som ikke alltid fremgår av katalogene eller variabeloversiktene. Noen variable som vanligvis følger standarder utarbeidet ved Statistisk sentralbyrå har også avvikende verdier. Disse avvikende verdiene må det tas hensyn til ved all tabellproduksjon.

### Statsborgerskap (stat)

Det er benyttet tallkode som korresponderer med Statistisk sentralbyrås standard (Statistisk sentralbyrå 1999, med referanse til Sentralkontoret for folkeregistrering 1996, eller [http://www.ssb.no/emner/00/00/20/nos\\_c513/](http://www.ssb.no/emner/00/00/20/nos_c513/)).

Følgende spesialkoder kan imidlertid forekomme:

Verdi	Betydning
00F	Norsk foretak
990	Ukjent
950	Ukjent, men utenlandsk
980	Statsløs

### Gjerningskommune (gjkom)

Det er benyttet kommunenummerering som korresponderer med Statistisk sentralbyrås standard (Statistisk sentralbyrå 1994). Følgende spesialkoder kan imidlertid forekomme:

Verdi	Betydning
2100	Svalbard
2111	Spitsbergen
2121	Bjørnøya
2131	Hopen
2200, 2211	Jan Mayen
2300	Kontinentalsokkelen
2311	Kontinentalsokkelen sør for 62° nord
2321	Kontinentalsokkelen nord 62° nord
??00	Fylkeskode der stedet ikke kan plasseres mer nøyaktig, 00 < ?? < 22
99??	Kode fra politiet: hele landet, ukjent

### Bokommune (bokom)

Det er benyttet kommunenummerering som korresponderer med Statistisk sentralbyrås standard (Statistisk sentralbyrå 1994). Følgende spesialkoder kan imidlertid forekomme:

Verdi	Betydning
2100	Svalbard
2111	Spitsbergen
2121	Bjørnøya
2131	Hopen
2200, 2211	Jan Mayen
2300	Kontinentalsokkelen
2311	Kontinentalsokkelen sør for 62° nord
2321	Kontinentalsokkelen nord 62° nord
99??	Kode fra politiet: hele landet, ukjent

## 7.5 Bruk av fellesressurser på UNIX

Under bearbeidingen benyttes noen ressurser som ligger sentralt på UNIX og er tilgjengelig for alle brukere. Endringer og fullstendig dokumentasjon av disse ressursene finnes sentralt og kan ikke inkluderes her. Det må undersøkes før hver årgang at ressursene tilsvarer det som ble brukt på tidligere årganger.

### Test av fødselsnummer

Fysisk fil: \$FELLES/sasprog/fnr sjekk.sas

SAS-program som tester et fødselsnummer med hensyn til matematisk gyldighet. Skiller mellom ugyldige fødselsnumre som har gyldig eller ikke gyldig fødselsdato. Testen slipper gjennom en lite del (ca 1 prosent) av de fødselsnumrene som har personnummer høyere enn 75000. Dette testes derfor i tillegg i SAS-programmene som benytter testen.

### Kommunekatalog

Fysisk fil: \$FELLES/ssbkat/kommune/g1994

Filbeskrivelse: \$FELLES/ssbkat/kommune/kommune.dok

### Postnummerkatalog

Fysisk fil: \$FELLES/ssbkat/poststed/g1994

Filbeskrivelse: -

Denne er midlertidig kopiert til \$KRIMINAL/felleskat/post.txt, siden fil under felleskatalogen ikke har vært til å stole på (endret både navn og variabelposisjoner i løpet av det siste året).

### Kommuneformater

Av felles formatene som finnes i SAS er \$kom94ur. benyttet

### Befolkningsfiler

Hvert år skal datamaterialet kobles til data fra Det sentrale folkeregisteret. Disse filene legges under felleskatalogen til kriminalstatistikken. Se rutine 06 under kapittel 3.2 for nærmere angivelse av filnavn og plassering. Seksjon for befolkning- og utdanningsstatistikk (320) må enten levere disse filene, eller det må gis tillatelse for å få lest nødvendige filer på dere områder. Uansett må seksjonen (og eventuelt kontor for IT, dvs. 303) kontaktes i god tid for at dette ikke skal forsinke produksjonen.

### Lydkoding

Fysisk fil: \$KRIMINAL/offer/prog/Soundex.sas

Rutine for lydkoding av tekststrenger er utviklet sentralt i Statistisk sentralbyrå. Det er imidlertid ikke opprettet en fast plassering for denne rutinen, slik det for eksempel er gjort med test av fødselsnummer. Foreløpig er rutinen kopiert til kriminalstatistikken område, se vedlegg 7.1

## De sist utgitte publikasjonene i serien Notater

- 1999/75 B. Halvorsen og M.I. Hansen: Dokumentasjon av utdrag fra skattestatistikken 1974-1994 for kobling mot forbruksundersøkelsen. 38s.
- 1999/76 H.P. Dahlslett og E. Engelian: Sentrumsstatistikk for Oslo og Akershus: Et pilotprosjekt. 50s.
- 1999/77 R. Eriksen og B. Haveråen: Inntekts- og formuesundersøkelsen for personlig næringsdrivende 1996: Dokumentasjon. 52s.
- 1999/78 P.M. Holt, S. Kristoffersen og V. Pedersen: Beregning av vektorer til inntekts- og formuesundersøkelsen 1997. 26s.
- 1999/79 P.M. Holt og T. Vevle: Skattestatistikk for rederier 1996 og 1997: Dokumentasjon. 26s.
- 1999/80 T. Bye, Ø. Døhl og J. Larsson: Klimagasskvoter i kraftintensive næringer. Konsekvenser for utslipp av klimagasser, produksjon og sysselsetting. Regionale konsekvenser. 11s.
- 1999/81 B. Mathisen: Flyktninger og arbeidsmarkedet 4. kvartal 1998. 39s.
- 1999/82 Ø. Kleven, E. Dalheim og D. Roll-Hansen: Innvandreres utdanning: - en pilotundersøkelse. 61s.
- 1999/83 E. Fidjestøl og I. Håland: Yrkeskatalog: Pr. desember 1999. 136s.
- 1999/84 T. Solberg: Virkning av revisjon på Avlingsstatistikk for jordbruksvekster i 1998. 24s.
- 1999/85 R. Choudhury, T. Eika og L. Haakonsen: KVARTS i praksis II: Systemer og rutiner i den daglige driften. 66s.
- 1999/86 G. Frøiland: Økonometrisk modellering av husholdningenes konsum i Norge: Demografi og formueseffekter. 55s.
- 1999/87 Y. Li: Beregning av elementær aggregater i konsumprisindeksen ved hjelp av generalisert gjennomsnitt. 41s.
- 1999/88 L. Rogstad og S.T. Vikan: Kobling av adresseregistrene i DSF og GAB 1999: Dokumentasjon av samsvar og avvik. 31s.
- 1999/89 E. Dalheim, J-A. S. Lie og D. Roll-Hansen: En skjemabasert komplettering av registeret over befolkningens høyeste utdanning - forprosjekt med fokus på innvandrere. 60s.
- 1999/90 K-A. Hovland og Å. Nossum: Flyreiser i konsumprisindeksen. 39s.
- 2000/1 E. Rønning: Utenlandske statsborgere og kommunestyrevalget 1999: Dokumentasjonsrapport. 34s.
- 2000/2 M. Bråthen: Personer registrert som yrkeshemmet i SOFA-søkerregisteret. 25s.
- 2000/3 A.K. Johnsen og Ø. Hokstad: FoB2001: Kvalitativ testing av boligskjema - prøveundersøkelse 1999: Dokumentasjonsnotat. 32s.
- 2000/4 C. Hendriks, Ø. Hokstad og R. Sønsterudbråten: FoB2001: Boligtelling - prøveundersøkelse 1999: Dokumentasjonsnotat. 60s.
- 2000/5 K. Bjønnes, G. Dahl og B.R. Joneid: FD - Trygd: Dokumentasjonsrapport: Økonomisk sosialhjelp 1992-1997. 31s.
- 2000/6 B.R. Joneid og J. Lajord: FD - Trygd: Dokumentasjonsrapport: Demografi 1992-1997. 117s.
- 2000/7 J. Heldal: Kalibrering av AKU: Dokumentasjon av metode og program. 28s.
- 2000/8 H. Hågård og L. Rogstad: FoB2001: Adresser i folkeregisteret og GAB: Rapport fra en arbeidsgruppe for adresse-samordning og utredning av elektronisk datautveksling mellom DSF og GAB. 51s.
- 2000/9 B. Sundby: Rutiner for produksjon av statistikk over pleie- og omsorgstjenestene i kommunene 1997. 84s.
- 2000/10 E. Aas: På leting etter målefeil - en studie av pleie- og omsorgssektoren. 31s.