

Interne notater

STATISTISK SENTRALBYRÅ

86/27

14. mai 1986

TEKNISK NOTAT OM OPPBYGGING AV MØDREFILEN.

EKTESKAPSEILEN OG KVINNEFILEN.

av

Øystein Kravdal

INNHOLD

	side
1. Innledning	1
2. Beskrivelse av 1984-versjonen av Mødrefilen	3
3. Prosedyre for oppbygging av ny versjon av Mødrefilen	5
4. Ressursbehov ved oppretting av ny Mødrefil	8
5. Omlegging av Beredskapsfilen	8
6. Utdrag av meldinger om endret ekteskapeleg status	10
7. Oppbygging av Ekteskapsfil	10
8. Beskrivelse av Ekteskapsfilen	13
9. Oppbygging av nye versjoner av Ekteskapsfilen	15
10. Oppretting av Kvinnefil ved kobling av Ekteskapsfil og Mødrefil	17
11. Beskrivelse av Kvinnefilen	19

1. INNLEDNING

Mødrefilen omfatter alle kvinner født i årene 1920 - 1971, og inneholder deres fødselshistorier samt en del kjennemerker fra Personregisterets Situasjonsfil pr. 31/12-1984. Ekteskapsfilen inneholder hovedsakelig ekteskapshistorier og en del opplysninger fra Folketellingen 1970. Den omfatter alle kvinner som har endret ekteskabelig status etter 1/10-1964 eller som var gift eller separert 1/11-1970. Kvinnefilen er dannet ved å slå sammen Mødrefilen og Ekteskapsfilen, og omfatter alle kvinner født 1920 - 1971, altså de samme som inngår i Kvinnefilen. Alle opplysninger i Mødrefilen og Ekteskapsfilen finnes også i Kvinnefilen.

Den eksisterende utgaven av Mødrefilen, heretter kalt 1984-versjonen, ble ferdig september 1985. Etter dette er det utført en kvalitetskontroll og et omfattende analysearbeid som er beskrevet i en rapport med forløpig tittel "En paritetsanalyse av norske kvinner på grunnlag av registerdata" samt i "Recent fertility trends in Norway" (presentert på Nordisk Demografisk Symposium, 13-16 juni 1986).

1984-versjonen har oppstått etter en unødvendig komplisert prosedyre. Dette skyldes at det underveis har kommet en rekke ideer om forbedringer. Fordi denne utgaven kan bli benyttet en del fremover før det lages en ny, vil vi gi en presis filbeskrivelse. En EDB-teknisk beskrivelse av oppbyggingen har imidlertid liten interesse.

Da Mødrefilen bør oppdateres årlig, presenterer vi et forslag til et enkelt opplegg som kan brukes i fremtiden.

Ekteskapsfilen ble opprettet i løpet av perioden oktober 1985 til januar 1986. Et notat som diskuterer filens kvalitet, er under utarbeiding av Helge Brunborg.

I januar 1986 ble Mødrefilen og Ekteskapsfilen koblet sammen til Kvinnefilen. Denne filen benyttes for øyeblikket i to prosjekter : "Barnetall i ekteskap 1980" og "Skilsmisseutviklingen i Norge". En nærmere omtale av datakvaliteten vil bli gitt i rapporter fra disse prosjektene.

Dette notatet er viet de rent tekniske aspekter ved oppbyggingen av Mødrefilen, Ekteskapsfilen og Kvinnefilen. Et kapittel omhandler også omleggingen av Beredskapsfilen. Det blir dessuten gitt detaljerte filbeskrivelser.

Systemarbeidet er gjort av Erling Henningstad (omlegging av Beredskapsfilen, oppbygging av Ekteskapsfilen) og Øystein Kravdal (_oppbygging av Mødrefilen og Kvinnefilen). Helge Brunborg har vært prosjektleder. Driftskontoret har gitt meget god assistanse i forbindelse med de mest ressurskrevende eksekveringene.

Noen av filene som er beskrevet i notatet, vil bli omdøpt fra P481 til PL81. Det er foreløpig uklart hvilke filer dette gjelder.

2. BESKRIVELSE AV 1984-VERSJONEN AV MØDREFILEN

1984-versjonen av Mødrefilen inneholder følgende kjennemerker for alle kvinner født mellom 1920 og 1971 som er tildelt norsk fødselsnummer :

	<u>Kjennemerke</u>	<u>Feltlengde</u>	<u>Posisjon</u>
1)	kvinnens fødselsnummer	11	1 - 11
1)	kvinnens ekteskapelige status	1	12 - 12
1)	dato for etablering av ekt.stat. (år,mnd.,dag)	6	13 - 18
1)	kvinnens registreringstype	1	19 - 19
1)	dato for etablering av reg. type (år,mnd.,dag)	6	20 - 25
1)	kvinnens bokommune	4	26 - 29
2)	antall levendefødte barn	2	30 - 31
3)	fødselsnummer for barn	11	
3)	kode for barn født i utlandet	1	
3)	fødselsnummeret til barnets far	11	
1)	Pr. 31/12-1984. (Situasjonsgifilen P441.S7220.A215U6A6.G85M2.V00 ble benyttet.)		
2)	Pr. 30/4-1985. (Også barn født i 1985 er altså med.)		
3)	Gjentas en gang for hvert barn (23 posisjoner pr. barn).		

Reg.type : 1 = Bosatt.
 2 = Død.
 3 = Utvandret.
 4 = Forsvunnet, dvs. formodet død (43 personer 31/12-1984).
 5 = Ikke bosatt, men med tilknytning til landet ved offentlig myndighet (593 personer 31/12-1984).
 6 = Dublett, dvs. vedkommende har annet fødselsnummer som skal brukes (21353 personer 31/12-1984).
 7 = Avgang, dvs. vedkommende har eventuelt annet fødselsnummer som skal brukes (2852 personer 31/12-1984).

- 4
- 8 = Dødfødt (262 personer 31/12-1984 ; svært få av de dødfødte får fødselsnummer).
- 9 = Korrigert fødselsnummer, dvs. vedkommende er tildelt nytt fødselsnummer etter retting i fødselsdato og/eller kjønn.
- blank = Tildelt fødselsnummer uten person, dvs. fødselsnummeret er tildelt utenom ordinær rutine, ingen person er ennå knyttet til nummeret.

I parentes er det oppgitt hvor mange personer i Personregisterets Situasjonstil pr. 31/12-1984 som står oppført med den aktuelle registreringstype. Disse opplysningene er gitt av Registerkontoret.

Ekteskapelig status : 1 = Ugift.
2 = Gift.
3 = Enke.
4 = Skilt.
5 = Separert.

Barn født i utlandet : 1 = Barnet er tildelt fødselsnummer ved innvandring etter 1/10-1964 (tildelingsgrunnlag 4 eller J).
0 = Ellers (dvs. barnet er tildelt fødselsnummer ved Personregisterets opprettelse 1/10-1964 eller ved fødselen; tildelingsgrunnlag ikke 4 og ikke J)

Filen omfatter 1514844 kvinner. Den er sortert etter kvinnes fødselsdato, men for hver kvinne ligger ikke fødselene i kronologisk rekkefølge.

Filnavnet er P481.S6932.I344A1A3.G85M6.V00. Filen er lagret på magnetbånd med variabel postlengde. Maksimal lengde er satt til $31 + 30 * 23 = 721$. Foreløpig har vi imidlertid ikke funnet noen kvinner som har mer enn 18 barn. Ved lagring på 3375-disker opptar filen 242 sylindere.

Ved bruk av variabel postlengde kan det oppstå problemer dersom postlengde og blokkstørrelse oppgis galt. Følgende DCB-statement ble brukt ved oppretting av Mødrefilen på magnetbånd :

DCB=(RECFM=VB,LRECL=721,BLKSIZE=32445).

3. PROSEDYRE FOR OPPBYGGING AV NY VERSJON AV MØDREFILEN

Det lar seg gjøre å benytte eksisterende versjon som utgangspunkt ved en oppdatering. Det vil da være nødvendig å korrigere registreringstype, ekteskapelig status etc. for alle kvinnene og addere de som har innvandret. Data om barna må oppdateres ved å påføre fødsler siste år samt de som er tildelt fødselsnummer ved innvandring. Det må dessuten korrigeres for eventuelle endringer av fødselsnummer ved hjelp av den såkalte Linkfilen. Dette ser ut til å bli en ganske komplisert rutine. Vi anbefaler derfor at Mødrefilen heller bygges opp helt fra grunnen hver gang det er behov for en ny versjon.

Sannsynligvis vil det være ønskelig å gjøre dette årlig, siden denne filen vil bli et viktig hjelpemiddel i analyse av blant annet fruktbarhet og familiestruktur. Det bør også bli aktuelt å bruke filen i statistikkproduksjon.

Mødrefilen er utelukkende basert på data fra Personregisterets Situasjonstilfil. Denne må gjennomleses to ganger. Mødrefilen bygges opp i fire trinn :

Trinn 1 :

Ved første gjennomlesing dannes en fil med disse kjennemerker for alle barn som har fått påført morens fødselsnummer og har reg.type 1,2 eller 3 :

barnets fødselsnummer,
morens fødselsnummer,
0/1-kode som viser om barnet er født i utlandet,
farens fødselsnummer,
reg.type for barnet,
dato for etablering av reg.type.

De to siste kjennemerkene er ikke med i den eksisterende utgaven av Mødrefilen.

I dette trinnet settes det ikke noen grense for hvilke kohorter av kvinner som skal være med.

For å få et inntrykk av datakvaliteten bør det ved gjennomlesing av Situasjonsfilen også telles opp for hvert fødselskull hvor mange barn som står oppført med mors fødselsnummer, og hvor mange som mangler dette. For hver av disse kategoriene bør det også bestemmes hvor mange barn som selv har fått tildelt fødselsnummer ved innvandring.

En slik gjennomlesing vil ta omtrent 5 CPU-minutter med den maskinkapasitet som for øyeblikket finnes på IBM-maskinen i Kongsvinger. Den totale tid som går med, vil imidlertid være betydelig lenger - selv hvis det er liten aktivitet på maskinen. Dette skyldes at Situasjonsfilen består av 10 magnetbånd.

Trinn 2 :

Neste trinn vil være å sortere filen fra trinn 1 etter morens fødselsnummer (dato + personnummer) og sekundært etter barnets, slik at det blir enkelt å føre opp fødslene i kronologisk rekkefølge.

Dette vil ta omtrent 10 CPU-minutter. Et oppsett for sortering finnes på filen RK81ØKR.COBOL.SOURCE(SYNKSX).

Trinn 3 :

Filen fra trinn 2 må ordnes slik at det bare blir en eneste post for hver kvinne. Denne må inneholde kvinnens fødselsnummer, antall barn og de data som er funnet for barna. Dette kan kalles en Mor-og-barn-fil, der alle mor-barn forhold fra Personregisteret (uansett morens fødselsår) har kommet med.

Denne Mor-og-barnfilen bør oppbevares i tillegg til den nye versjonen av Mødrefilen. Den kan nemlig være en aktuell datakilde ved studier av periodedata for årene etter 1953. De fleste fødsler i disse årene vil nemlig finnes, men pariteten kan være gal, særlig for kvinner født før 1935.

Det kan også være aktuelt å foreta en del kontroller på dette trinnet, for eksempel om moren er mellom 12 og 50 år når hun får barnet, eller om siste barnet er født før hun eventuelt dør eller utvandrer.

Det må tas stilling til om slike feil skal føre til at posten forkastes (i så fall er det viktig å passe på at kvinnen ikke kommer inn som barnløs i trinn 4) eller om det skal gjøres rettelser i posten. Alternativt kan feilene bare telles opp.

Trinn 3 vil sannsynligvis ta omtrent 5 CPU-minutter.

Trinn 4 :

I fjerde trinn kobles det til Situasjonsfilen slik at kvinner som ikke har fått barn, kommer med. Det må bestemmes hvilke kohorter som skal inkluderes i den endelige filen. I denne fasen tas det for alle kvinner også med opplysninger om :

ekteskapelig status,
dato for etablering av ekteskapelig status,
registreringstype,
dato for etablering av registreringstype,
bostedskommune,

som i 1984-utgaven. I tillegg bør følgende kjennemerker inkluderes :

tildelingsgrunnlag og tildelingsdato for fødselsnummer,
statsborgerskap,
ektefelles fødselsnummer.

Det kan kanskje også være aktuelt å ta med foreldrenes fødselsnummer for å ha mulighet til å utføre tre-generasjons-studier.

Den filen som dannes i trinn 4, og som er den nye versjonen av Mødrefilen, inneholder dermed alle kvinner født i et visst tidsrom som har fått fødselsnummer (uansett registreringstype).

I trinn 4 bør det også foretas en enkel optelling av antall fødsler. Dette kan gi en pekepinn om eventuelle feil under oppbyggingen av filen.

For eksempel er det meget enkelt å finne antall kvinner bosatt i landet ved utgangen av det siste året filen omfatter, og deres barnetall (0,1,2,3,4+) på dette tidspunkt. En slik opptelling bør gjøres for hver kohort separat.

Dette fjerde trinnet vil ta omtrent 10 CPU-minutter.

Programmene kan lages raskt ved å modifisere programmet RK81ØKR.COBOL.SOURCE(MERGE) som styres fra (MERGEX).

4. RESSURSBEHOV VED OPPRETNING AV NY MØDREFIL

De nødvendige EDB-program i forbindelse med oppretting av en ny versjon vil være svært enkle å lage, i og med at vi allerede har høstet en del erfaringer som kan være av stor verdi. Det er neppe nødvendig med mer enn 100 systemtimer i 1986 for å lage en ny versjon.

Vi foreslår at det legges ned et visst arbeid i å lage vedlikeholdsvennlige og godt dokumenterte programmer, slik at disse kan benyttes så godt som uendret senere år. Fra og med 1987 vil det kanskje være tiltrekkelig med 50 systemtimer årlig for å få nye versjoner av Mødrefilen.

Ved analyse av individdata kan det ofte bli svært tidkrevende programmer. Det kan derfor være ønskelig å benytte en miniversjon av Mødrefilen - i det minste i de innledende faser av et forskningsprosjekt. Derfor bør det kanskje lages en fil der for eksempel bare hver femte kohort er med. Oppsett for dette finnes i RK81ØKR.COBOL.SOURCE(RANDKV) og (RANDKVX).

Oppsett for å overføre fra disk til magnetbånd finnes på filen RK81ØKR.COBOL.SOURCE(TAPE), og oppsett for å kopiere fra magnetbånd til disk på RK81ØKR.COBOL.SOURCE(DISK).

5. OMLEGGING AV BEREDSKAPSFIL

Etter at Personregisteret ble opprettet 1/10-1964, har meldinger om endringer av

viktige kjennemerker, som for eksempel ekteskapelig status, blitt lagret på en fil kalt Beredskapsfilen. Filen inneholder både såkalte Historiekstrakt-poster og Statistikkstrakt-poster. Alle meldinger om endret ekteskapelig status er av førstnevnte type. I disse meldingene fremgår det hvilke kjennemerker som sto i Personregisteret før den nye meldingen kom, hvilken dato denne meldingen kom, og hva slags meldingstype det var.

En versjon av Beredskapsfilen som inneholder alle meldinger for perioden 1/10-1964 til 30/4-1985 ble i juni 1985 overført til IBM-maskinen på Kongsvinger.

Den var av praktiske grunner lagret på to forskjellige filer. På den ene var det meldinger for årene 1964 - 1970 og på den andre meldinger for 1980 - 1985. Filnavnene var :

PL41.S6908.A003A6A6.G6479.V01,
PL41.S6908.A003A6A6.G8085.V01.

I løpet av perioden juni til september 1985 ble det utført et betydelig arbeid for å gjøre Beredskapsfilen mer tilgjengelig for forskningsformål (se "Omlegging av beredskapsfilen", av Helge Brunborg, Interne notater 85/41).

Første trinn i omleggingen var å koble hver av de to filene som utgjør Beredskapsfilen, med den såkalte Linkfilen (PL41.S7220.A226A5A2.G85M4.V01). Samtidig ble det koblet med en fil som viste forbindelsen mellom de gamle meldingstypene og de nye som skulle innføres for å forenkle bruken av Beredskapsfilen. Resultatet av denne koblingen var at det ble dannet ett par med filer for 1964-1979 og et annet par for 1980-1985. Den ene filen i paret inneholdt meldinger der fødselsnummeret var endret (ifølge opplysningene i Linkfilen), mens den andre inneholdt meldinger der fødselsnummeret ikke var endret. Den siste var selvfølgelig langt større enn den første. Koblingen ble utført ved hjelp av programmene B457PA00 for 1964-1979 og B458PA00 for 1980-1985.

I neste trinn ble de to filene der fødselsnumrene var endret, sortert etter det nye fødselsnummeret. Deretter ble den store og den lille filen i hvert par sammenkoblet. Dette ble gjort ved hjelp av programmet B460PA00.

Etter disse koblingene besto Beredskapsfilen av bare en fil for 1964-1979 og en for 1980-1985. Disse to ble så koblet sammen ved programmet B460PA00. Deretter ble det fjernet dubletter (omtrent 7000) ved programmet B478PA00. Den resulterende filen er :

PL41.S6908.A003B1A1.G6485.V00.

Den består av 12394895 poster (en post for hver melding). Postene er sortert slik at alle meldinger for samme individ kommer etter hverandre i kronologisk rekkefølge. Individene er ordnet etter fødselsnummer.

6. UTDRAG AV MEDLINGER OM ENDRET EKTESKAPELIG STATUS

Under oppbygging av Ekteskapsfilen var det bare et utvalg av meldingene i Beredskapsfilen som var interessante (melding 211 - 383, 421 - 431 og 911 - 912). Disse ble trukket ut ved programmet B513PA00. Resultatet ble lagt på filen :

P481.S6932.I381A1A1.G6485.V00.

Denne inneholder omtrent 2 millioner meldinger for både menn og kvinner og vil danne utgangspunktet ved en eventuell oppbygging av en Ekteskapsfil for menn.

En fil med meldinger for bare kvinner finnes på :

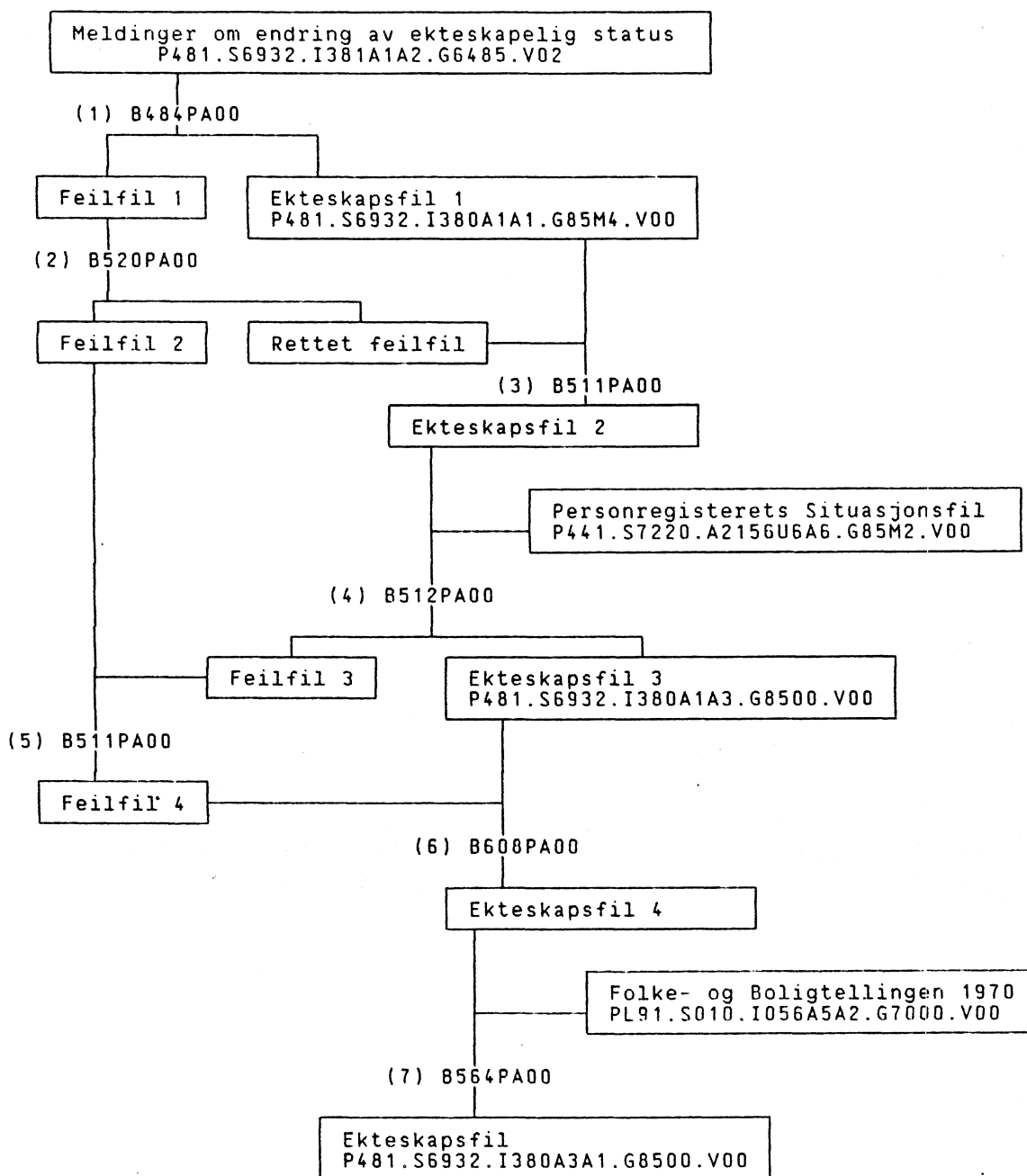
P481.S6932.I381A1A2.G6485.V02

Denne filen er brukt til å lage kvinners ekteskapshistorier.

7. OPPBYGGING AV EKTESKAPSFIL

Oppbyggingen er gjort i 7 trinn. Disse beskrives skjematisk ved figuren på neste side. For filer som fortsatt eksisterer, er navnet oppgitt.

Den endelige filen kalles Ekteskapsfil. Tidligere versjoner kalles Ekteskapsfil 1, Ekteskapsfil 2, Ekteskapsfil 3 og Ekteskapsfil 4.



- (1) På dette trinnet ble alle ekteskapsmeldinger for samme person lagt på samme post. Posten har en komponent med fast lengde og en med variabel. En teller angir hvor mange meldinger den variable delen består av. Slik denne posten er organisert, inneholder hver melding den tidligere ekteskapelige status, dato for endring av status, meldingstype og eventuelt tidligere ektefelles fødselsnummer. Det sistnevnte feltet inneholder ingen informasjon med mindre meldingen gjelder et gjengifte. På dette trinnet ble det dannet en post for alle individer som endret ekteskapelig status mellom 1/10-1964 og 30/4-1985. Meldinger som gjelder 1985 ble imidlertid strøket, slik at en del

poster inneholder 0 meldinger. Under opprettingen av en slik fil med ekteskapshistorier ble det utført en del kontroller. Det ble undersøkt om overgangen fra en status til en annen var lovlig (feil av type A) og om det var samsvar mellom meldingstype og tidligere status (feil av type B). Tabeller som viser forekomsten av feil etter status, medlingstype og dato finnes i Brunborgs notat, som er under utarbeiding. I alt ble det funnet omtrent 16000 historier med inkonsistens. Disse ble lagt på en egen fil. Det var omtrent 780000 historier der det ikke var inkonsistens.

- (2) Programmet leser filen med 16000 inkonsistente historier og vurderer om det kan gjøres opprettinger. Det viste seg at 3000 historier ble akseptable ved å fjerne dubletter. Den største del av de gjenværende historiene hadde feil i første melding. Sannsynligvis har det kommet inn feil ved oppretting av Personregisteret. Det ble derfor utført en del substitusjoner av ekteskapeleg status. Dette ble tydelig markert på den resulterende filen (se filbeskrivelse i kapittel 8). Etter dette trinnet besto feilfilen (Feilfil 2) av omtrent 1200 poster.
- (3) Filen med rettede historier (omtrent 14800) ble koblet med ekteskapshistorien fra første trinn.
- (4) Det kobles med Personregisterets Situasjonfil for å finne ekteskapeleg status den 31/12-1984 og ektefelles fødselsnummer. Disse opplysningene ble lagt inn som en siste melding med dato 1/1-1985 og meldingstype 884. Dette innebærer at de aller fleste poster har minst to meldinger. I de tilfeller hvor kvinnen bare har hatt endring i ekteskapeleg status etter 31/12-1984 derimot, finnes bare meldingen som viser data fra Situasjonfilen. Under denne koblingen ble det foretatt en kontroll der status den 31/12-1984 ble sammenliknet med status og meldingstype på en eventuell tidligere melding. Det ble også kontrollert om kjennemerket "dato for siste endring av ekteskapeleg status" stemte overens med ekteskapshistorien. Det ble funnet omtrent 2800 og 400 feil i disse testene. Historier som ble rammet av den førstnevnte kontrollen, ble lagt ut på egen fil (Feilfil 3).
- (5) De to feilfilene, med omtrent 1200 (Feilfil 2) og 2800 (Feilfil 3) poster, ble koblet sammen til en fil med 4000 poster (Feilfil 4).
- (6) Den resulterende Feilfil 4 ble koblet sammen med ekteskapshistorien fra

trinn 3. Historier som inneholdt inkonsistenser (dvs. poster fra Feilfil 4), ble merket med en egen kode (se filbeskrivelsen i kapittel 8). Det er her viktig å merke seg at de postene som kommer fra Feilfil 2, ikke inneholder data fra Situasjonsfilen. Disse postene inneholder forøvrig minst en melding fra perioden 1/10-1964 til 31/12-1984.

- (7) Folke- og boligtellingsdata ble gjennomlest, og for alle som var gift eller separert (929443 individer), ble kjennemerkene "ekteskapsår" og "antall barn i nåværende ekteskap" ekstrahert ut. Dette ekstraktet ble så sortert og koblet til ekteskapshistorien fra foregående trinn (793777 individer). Det ble laget en ny fil som inneholdt alle individer som var med i minst en av de to filene. Denne inneholder 1340915 individer. Programmet i dette trinnet er skrevet i EASYTRIEVE, alle de andre i COBOL.

8. BESKRIVELSE AV EKTESKAPSFILLEN

	<u>Kjennemerke</u>	<u>Feltlengde</u>	<u>Posisjon</u>
	kvinnens fødselsnummer	11	1 - 11
1)	posttype	1	12 - 12
2)	feilkode	1	13 - 13
	ekteskapsår oppgitt i FOB 1970	2	14 - 15
	antall levendefødte barn oppgitt i FOB 1970	2	16 - 17
3)	antall meldinger om endret ekteskapeleg status	2	18 - 19
4) 5)	korreksjon av ekteskapeleg status	1	
4) 6)	ekteskapeleg status	1	
4)	dato for endring av ekteskapeleg status	6	
4)	meldingsnummer	3	
4)	ektefelles fødselsnummer	11	

Forklaringer :

Dato er gitt som år, måned, dag.

1) Posttype :

0 : Bidrag fra FOB 1970 + Ekteskapsfil 4 (fra trinn 6)

1 : Bidrag fra Ekteskapsfil 4

2 : Bidrag fra FOB 1970

Hvis posttype er 2, er det 0 meldinger.

2) Feilkode (se trinn 6) :

1 : Har funnet inkonsistens i ekteskaphistorien, men har ikke foretatt substituering, dvs. posten kommer fra Feilfil 4.

0 : Ellers (dvs. enten er det ikke funnet inkonsistens, eller det er funnet inkonsistens, men foretatt substituering).

3) Siste melding gjelder status 31/12-1984 (unntatt noen få tilfeller der feilkode er 1). Ingen meldinger for 1985 er med.

4) Disse felt blir gjentatt 0 - 20 ganger hver avhengig av antall meldinger (posisjon 58 - 59). Det er sortert etter dato for begivenheten.

5) Korreksjon av ekteskapelig status (se trinn 2) :

1 : Substituert ekteskapelig status på grunn av antatt feil.

0 : Ellers.

6) Ekteskapelig status :

1 : Ugift.

2 : Gift.

3 : Enke.

4 : Skilt.

5 : Separert.

9. OPPBYGGING AV NY VERSJON AV EKTESKAPSFIL

Det kan være aktuelt å bygge opp en tilsvarende Ekteskapsfil for menn om forholdsvis kort tid. Det er da rimelig å ta utgangspunkt i de ekteskapsmeldingene som er plukket ut fra Beredskapsfilen, og legge de meldingene som angår menn, på en egen fil. Deretter kan de eksisterende programmene benyttes med meget små modifikasjoner. Sannsynligvis er det bare nødvendig å endre noen få tester som omfatter meldingstype.

Visse forenklinger er mulige i forhold til den oppdelingen i 7 trinn som er forklart tidligere. For eksempel kan trinn 1, 2 og 3 slås sammen ved å foreta substitusjoner og fjerning av dubletter ved første gjennomgang av ekstraktet fra Beredskapsfilen. Dessuten blir trinn 5 og 6 overflødig hvis poster med inkonsistenser ikke legges ut på egen fil, men bare markeres med en egen kode etter hvert som de oppdages. Oppbyggingen vil da bare ha 3 trinn :

- (a) Oppbygging av ekteskapshistorier på grunnlag av meldinger fra Beredskapsfilen.
- (b) Kobling med Situasjonsfilen.
- (c) Kobling med FOB 1970.

Det er grunn til å anta at oppbygging av en fil med ekteskapshistorier for menn ikke vil kreve mer enn 100 systemtimer (iberegnet den tid som går med til å utføre eksekveringene). Trolig vil det gå med noe mer tid hvis programmene skal bygges om til en tre-trinns prosess, men neppe mer enn 50 timer ekstra. En slik ombygging vil være tidsbesparende på lengre sikt, da det er ønskelig at det etableres nye Ekteskapsfiler regelmessig.

Beredskapsfilens skjebne er i skrivende stund uklar, men det er i alle fall sterke ønsker fra Sosiodemografisk forskningsgruppe om at historiske individopplysninger må lagres på en eller annen måte. Nedenfor er det antatt at en ny Beredskapsfil med data for 1/5-1985 og senere vil være tilgjengelig.

Ved oppbygging av nye Ekteskapsfiler med individhistorier ut over 30/4-1985 vil nok det gunstigste være å ta utgangspunkt i Beredskapsfilen. Den siste versjonen (PL41.S6908.A00381A1.G6485.V00) må kobles sammen med en ny Beredskapsfil med meldinger fra og med 1/5-1985. I denne koblingen må det tas hensyn til eventuelle forskjeller i koder og filorgansiering mellom de to filene. Linkfilen må også benyttes. Deretter trekkes de aktuelle meldinger om

endring av ekteskapeleg status ut, og trinnene a,b,c kan så utføres.

I trinn a dannes det en fil som inneholder ekteskaphistoriene til alle som har skiftet status i årene 1965 og senere. Det vil imidlertid ikke finnes opplysninger om fødselsnummeret til siste ektefelle og heller ikke siste registrerte ekteskapelege status. Dette kan fås ved å koble med Situasjonsfilen (trinn b). For mange formål kan det også være nødvendig med opplysninger om de som ikke har endret status etter 1965. Av denne grunn bør koblingen med Situasjonsfilen utføres på en slik måte at alle som har fått tildelt fødselsnummer, er med i den resulterende filen.

Trinn b kan imidlertid sløyfes helt dersom en ny utgave av Ekteskapsfilen ikke skal brukes alene, men alltid sammenkoblet med en tilsvarende ny utgave av Mødrefilen (eller eventuelt en Fedrefil). Dette skyldes at kommende versjoner av Mødrefilen sannsynligvis vil inneholde ektefelles fødselsnummer og ekteskapeleg status. Det er nemlig enkelt å få med disse kjennemerkene i Mødrefilen, da det under etablering av den likevel skal utføres en kobling med Situasjonsfilen. Et annet poeng som må nevnes i denne sammenheng, er at det ved nuptialitetsanalyser kan være nødvendig å kjenne dato for eventuell død eller utvandring. Også dette kan fås ved kobling med Mødrefilen. Dersom Ekteskapsfilen skal brukes alene, må dette kjennemerket inkluderes i trinn b.

Trinn c bør kanskje utvides noe, slik at det for alle individer tas med en del folketellingsopplysninger, ikke bare for de som er gift eller separert på tellingstidspunktet. For de gifte og separerte bør barnas fødselsår tas med. Det må imidlertid understrekes at det finnes gode grunner til å ikke ta med folketellingsopplysninger i det hele tatt.

Etter all sannsynlighet vil det ikke være nødvendig med mer enn 200 systemtimer for å få konstruert en ny Ekteskapsfil. Det kan være aktuelt å få gjort dette forholdsvis regelmessig, for eksempel hvert tredje eller hvert femte år. Mødrefilen derimot bør lages årlig, eller i allefall ikke med mer enn tre års mellomrom. Det er viktig at det ved konstruksjon av en ny Ekteskapsfil lages en Mødrefil som er oppdatert frem til samme dato.

10. OPPBYGGING AV KVINNEFIL VED KOBLING AV EKTESKAPSFIL OG MØDREFIL

I januar 1986 ble Ekteskapsfilen og en utvidet versjon av Mødrefilen (begge ajourført til 31/12-1984) koblet sammen. Utvikling av programmer og forberedelse til eksekvering gikk forholdsvis raskt. Det ble bare brukt omtrent 40 arbeidstimer på dette.

1984-versjonen av Mødrefilen (P481.S6932.I344A1A3.G85M6.V00) inneholder fødselshistorier for alle kvinner født 1920-71. For at Kvinnefilen skulle inneholde ektefelles fødselsnummer også for personer som ikke har hatt endring i ekteskapelig status etter 1/10-1965 (det vil si kvinner som ikke var med i Ekteskapsfilen ved utgangen av trinn 3), ble den utvidet ved å koble mot Situasjonssfilen (ajour 31/12-1984). I denne koblingen ble det tatt med :

ektefelles fødselsnummer,
tildelingsgrunnlag og tildelingsdato for fødselsnummer.

Koblingsprogrammet ligger på filen RK81ØKR.COBOL.SOURCE(EKTPNR), som styres fra (EKTPNRX). Den utvidede Mødrefilen ligger på P481.S6932.I344A3A1.G85M6.V00. En nøyaktig filbeskrivelse finnes innenfor IBM-dokumentasjons-systemet og blir ikke gjengitt i dette notatet.

Mens den utvidede Mødrefilen inneholder alle individer født 1920-1971, men ingen født tidligere, inneholder Ekteskapsfilen (P481.S6932.I380A3A1.G8500.V00) alle personer som var gift eller separert den 1/11-1970 eller som har hatt endring av ekteskapelig status i perioden 1/10-1964 til 30/4-1985. Ved kobling tok vi bare med kvinner som representert i Mødrefilen. Det vil si at vi fikk en Kvinnefil med fødsels- og ekteskapshistorie for alle kvinner født 1920-1971. I prosjektet "Barnetall i ekteskap 1980" har dette vist seg å være en svakhet. Der er det nemlig også av betydning å kjenne ekteskapshistoriene til eldre kvinner og dessuten deres eventuelle døds- eller utvandningsdato. Det er derfor opprettet en fil med en del data om kvinner født før 1920. Denne har navnet PL81.S6932.I426A1A3.G8500.V00 og har samme format som Kvinnefilen (kjennemerker som mangler er satt lik 0 eller blank).

Programmet RK81ØKR.COBOL.SOURCE(EKTKVI), som styres fra (EKTKVIX), koblet den utvidede Mødrefilen og Ekteskapsfilen sammen til Kvinnefilen :

P481.S6932.I426A1A1.G8500.V00.

Denne filen består av to bånd, og beslaglegger omtrent 400 sylindere på 3375-disker.

Fordi Kvinnefilen hadde to variable komponenter og skulle leses av et FORTRAN-program, ble det lagt inn en del blanke felt under koblingen. Fødselsåret ble dessuten tatt ut av fødselsnummeret som eget kjennemerke for å forenkle innlesingsmodulene i analyseprogrammene.

Programmet RK81ØKR.COBOL.SOURCE(MINI), styrt av (MINIX), lager en versjon der bare enkelte kohorter plukkes ut. Kohortene 1935,1940,1945,1950,1955 og 1960 er på denne måten tatt ut og lagt på eget bånd : P481.S6932.I426A1A2.G8500.V00.

11. BESKRIVELSE AV KVINNEFILEN

	<u>Kjennemerke</u>	<u>Feltlengde</u>	<u>Posisjon</u>
1)	kvinnens fødselsnummer	11	1 - 11
1) 2)	kvinnens ekteskapelige status	1	12 - 12
	dato for etablering av ekteskapelig status	6	13 - 18
3)	ektefelles fødselsår	2	19 - 20
3)	ektefelles fødselsnummer	11	21 - 31
4)	grunnlag for tildeling av fødselsnummer	1	32 - 32
	dato for tildeling av fødselsnummer	6	33 - 38
5)	kvinnens registreringstype	1	39 - 39
	dato for etablering av registreringstype	6	40 - 45
1)	kvinnens bokommune	4	46 - 49
6)	antall levendefødte barn	2	50 - 51
7)	posttype	1	52 - 52
8)	feilkode	1	53 - 53
	ekteskapsår oppgitt FOB 1970	2	54 - 55
	antall levendefødte barn oppgitt FOB 1970	2	56 - 57
9)	antall meldinger om endret ekteskapelig status	2	58 - 59
10)	blank	2	
10)	fødselsnummer for barn	11	
10) 11)	kode for barn født i utlandet	1	
10)	fødselsnummeret til barnets far	11	
12) 13)	korreksjon av ekteskapelig status	1	
12) 2)	ekteskapelig status	1	
12)	dato for endring av ekteskapelig status	6	
12)	meldingsnummer	3	
12)	ektefelles fødselsår	2	
12)	blank	1	
12)	ektefelles fødselsnummer	11	

Forklaringer :

Dato er gitt ved år, måned, dag.

1) Pr. 31/12-1984.

2) Ekteskapelig status :

1 : Ugift.

2 : Gift.

3 : Enke.

4 : Skilt.

5 : Separert.

3) Eventuell ektefelle pr. 31/12-1984

4) Tildelingsgrunnlag :

blank : Bosatt pr. 1/10-1964.

0 : Bosatt pr. 1/10-1964.

4 : Innvandring.

5 : Fødsel.

6 : Skapkort.

7 : Mangelskjema. Tilgang manko, ny.

9 : Tilgang manko, gammel.

S : Endelig utført fødselsnummerkorleksjon og ren tilgang.

J : Innvandring, tildelt som ren tilgang.

5) Registreringstype :

1 : Bosatt.

2 : Død.

3 : Utvandret.

4 : Forsvunnet, dvs. formodet død (43 personer 31/12-1984).

5 : Ikke bosatt, men med tilknytning til landet ved offentlig myndighet (593 personer 31/12-1984).

6 : Dublett, dvs. vedkommende har annet fødselsnummer som skal brukes (21353 personer 31/12-1984).

7 : Avgang, dvs. vedkommende har eventuelt annet

- fødselsnummer som skal brukes (2852 personer 31/12-1984).
- 8 : Dødfødt (262 personer 31/12-1984 ; svært få av de dødfødte får fødselsnummer).
- 9 : Korrigert fødselsnummer, dvs. vedkommende er tildelt nytt fødselsnummer etter retting i fødselsdato og/eller kjønn.
- blank : Tildelt fødselsnummer uten person, dvs. fødselsnummeret er tildelt utenom ordinær rutine, ingen person er ennå knyttet til nummeret.

I parentes er det oppgitt hvor mange personer i Personregisterets Situasjonstil pr. 31/12-1984 som står oppført med den aktuelle registreringstype. Disse opplysningene er gitt av Registerkontoret.

- 6) Pr. 30/4-1985. Også barn født i 1985 er altså med.
- 7) Posttype :
- 0 : Bidrag fra Mødrefil + FOB 1970 + Ekteskapsfil 4 (fra trinn 6)
 - 1 : Bidrag fra Mødrefil + Ekteskapsfil 4
 - 2 : Bidrag fra Mødrefil + FOB 1970
 - 3 : Bidrag bare fra Mødrefil
- 8) Feilkode :
- 1 : Har funnet inkonsistens i ekteskaphistorien, men har ikke foretatt substitusjon.
 - 0 : Ellers
- 9) Siste melding gjelder status 31/12-1984 (unntatt noen få tilfeller der feilkode er 1). Ingen meldinger for 1985 er med.
- 10) Gjentas 0 - 30 ganger avhengig av antall barn (posisjon 50 - 51). Det er sortert etter barnas fødselsdato.
- 11) Kode for barn født i utlandet :
- 1 : Tildelingsgrunnlag 4 eller J, dvs. født i utlandet.
 - 0 : Ellers.

12) Disse felt blir gjentatt 0 - 20 ganger hver avhengig av antall meldinger (posisjon 58 - 59). Det er sortert etter dato for begivenheten.

13) Korreksjon av ekteskapeleg status :

1 : Substituert ekteskapeleg status på grunn av antatt feil.

0 : Ellers.