



Torill Vangen

Notater

**Nasjonal utdanningsdatabase
NUDB**
Dokumentasjonsrapport

Datavarehus for utdanningsdata,
1970-2006

Forord

Dette notatet dokumenterer forløpsdatabasen Nasjonal utdanningsdatabase (heretter kalt NUDB). I tillegg til at dette notatet beskriver hvilke data som inngår i NUDB, inneholder det også en dokumentasjon av de arbeidsmetoder som er benyttet og den statistikkfaglige vurderingen som er lagt til grunn for å omforme og tilrettelegge disse data for innlasting i NUDB. Mye arbeid har vært lagt ned i å gjøre dataene sammenlignbare over tid slik at det skal være mulig å ta sammenhengende forløpsdata ut av NUDB, også der hvor det på grunnlagsfilene er benyttet forskjellige kodelister over tid.

For brukere av data fra NUDB vil kap.8. Feil og mangler i dataene. Endringer i prinsipper, definisjoner og avgrensing av populasjonen, gi viktig informasjon.

Dokumentasjonen i dette notatet omhandler bl.a. data fra filene for høyeste fullførte utdanning, igangværende/avsluttet utdanning pluss diverse befolkningsfiler. Filene inneholder utdanningsopplysninger for hele befolkningen for årene fra og med 1970 og fram til i dag.

I arbeidet med å tilrettelegge data for NUDB er det mange personer som har bistått prosjektgruppa med faglig råd og veiledning. Vi nyter her anledningen til å takke medlemmene i referansegruppa for NUDB. Disse er Per Olaf Aamodt fra Norsk institutt for studier av forskning og utdanning(NIFU), Oddbjørn Raaum fra Frischsenteret, Jonny Einarsen fra Læringscenteret og Morten Nordlie, Odd Larsen, Mads Gravås, Dankert Vedeler, Erik Dahl og Marie Arneberg fra Utdannings- og forskningsdepartementet.

En spesiell takk går også til Jorunn Lajord for verdifull bistand i forbindelse med spesifikasjon av data.

Notatet er blitt til ved å samle dokumentasjon som er skrevet underveis av de ulike medlemmene av prosjektgruppen, samt dokumentasjon som grunnlag for kap. 8 fra Tor Jørgensen. Torill Vangen har hatt redigeringsansvaret for notatet.

Dokumentasjon av dataene i NUDB, med variabler og kodelister finnes på <http://www.ssb.no/mikrodata/>

Innhold

1. Introduksjon til NUDB	7
1.1. Bakgrunn for modellvalg.....	7
1.2. Definisjon av en forløpsdatabase	7
1.2.1. Hva er en forløpsdatabase?.....	7
1.2.2. Revisjon av forløpsdata	8
1.3. Fordeler og ulemper ved den valgte modellen	8
1.4. Resultat.....	8
1.5. Arbeidsflyten i NUDB prosjektet.....	9
2. Tilrettelegging av grunnlagsregistrene.....	11
2.1. Beskrivelse av grunnlagsregistre.....	11
2.1.1. Filidenter	11
2.1.1.1. Igangværende utdanning per 1.okt.:	11
2.1.1.2. Avslutta utdanning i "skole"-året	12
2.1.1.3. Befolkningens høyeste utdanning (BHU) per 1.okt.	12
2.1.2. Filbeskrivelser for grunnlagsregistrene	13
2.1.2.1. Igangværende utdanning per 1.okt. og avslutta utdanning i løpet av året	13
2.1.2.2. Befolkningens høyeste utdanning per 1.oktober	18
2.2. Korreksjoner og kontroller	22
2.2.1. Fødselsnummerkontroll.....	22
2.2.2. Dublettkontroller	22
2.2.2.1. Igangværendefilene:.....	22
2.2.2.2. Avsluttafilene:.....	22
2.2.2.3. BHU-filene:.....	23
2.2.3. Datokontroll.....	23
2.2.3.1. Igangværende- og avsluttafilene:	23
2.2.3.2. BHU-filene:.....	24
2.2.4. Endring og korreksjon på variabler	25
2.2.4.1. Igangværendefilene:.....	25
2.2.4.2. Avsluttafilene:.....	28
2.2.4.3. BHU-filene:.....	28
3. Tiltrettelegging av miniregistrene.....	32
3.1. Variable på miniregistre	32
3.1.1. Igangværende- og avslutta utdanning.....	32
3.1.2. BHU.....	34
3.2. Korreksjoner og kontroller	36
3.2.1. Kontroller på filene for igangværende utdanning per 1.okt. hvert år	36
3.2.2. Kontroller på filene for avslutta utdanning i skoleåret.	41
3.2.3. Kontroller på filer for befolkningens høyeste utdanning, per 1.okt. hvert år	49
3.3. Avvik fra tidligere BHU-publisering.....	54
4. Diagrammer for miniregisterrutiner	55
5. Tilrettelegging av forløp i NUDB	69
5.1. Kursfilen. Forløp. (F_UTD_KURS).....	69
5.2. Viktigste kurs. Personnivået. Forløp. (F_UTD_PERSON).....	69
5.2.1. Generelt	69
5.2.2. STUDIEKODE på personnivået	70
5.2.3. Regelverk for valg av viktigste kurs.....	72

5.2.4.	Regler for telling av "antall semester innen studiet" ved fullføring av utdanning/grad.....	73
5.2.5.	Regler for å telle normert tid for de som har fullført en utdanning	73
5.2.5.1.	Videregående utdanning:	73
5.2.5.2.	Høyere utdanning:.....	73
5.3.	Demografitabellen, daterte variable, årlig datering.....	76
5.3.1.	Høyeste fullførte utdanning(BHU).....	76
5.3.2.	BU. Ny gruppering av befolkningens utdanningsnivå	78
6.	Tabellene i NUDB med variabeloversikter	81
6.1.	Demografitabellen. Demografiske opplysninger samt høyeste fullførte utdanning(BHU og BU, gammel og ny definisjon). Forløp med årlig datering per 1. oktober.....	81
6.2.	Kurstabellen. Forløp med månedlig datering	81
6.3.	Viktigste kurs. Forløp med månedlig datering	83
6.4.	Persontabellen, faste variable på personnivå. (ikke forløp).....	83
6.5.	Karakterer for avsluttet grunnskole.....	85
6.6.	Vitnemål fra videregående utdanning. Data fra Nasjonal vitnemålsdatabase(NVB)	86
6.7.	Fag fra videregående utdanning. Data fra Nasjonal vitnemålsdatabase(NVB).....	87
6.8.	Andre filer ifm NUDB	88
7.	Statistikk- og IT-faglige vurderinger	89
7.1.	Mulige konsekvenser av påføring av vektall på eldre filer	89
7.2.	Foreldrenes høyeste utdanning/sosial bakgrunn.....	90
7.2.1.	Foreldrenes utdanning settes en gang	90
7.2.2.	Gruppering av foreldrenes utdanningsnivå.....	91
7.2.3.	Bostedskommune	91
7.2.4.	Bostedskommune ved 16-års alder eller "hjemstedskommune"	92
7.3.	Hoved (hovedgruppe).....	92
7.4.	Statsborgerskap	92
7.5.	Skoleslag	92
7.5.1.	SSLAG	92
7.5.2.	UTD.....	93
7.6.	Skoleidentifikasjon.....	94
7.7.	"Manglende vårsemester"	94
7.8.	Ktrinn (Klassestrinn gruppert. Gjelder utdanninger under Opplæringsloven/Lov om videregående opplæring/Lov om fagskoler).....	94
7.9.	Normert tid i utdanning	95
7.9.1.	Telling av antall semester i NUDB.....	96
7.10.	Definisjon av semester	97
7.11.	Gjennomstrømningsvariable i NUDB	97
7.11.1.	Gjennomstrømningsvariable som settes en gang/ved 1. gangs hendelse. (TAB_UTD_PERSON).....	97
7.11.2.	Gjennomstrømningsvariable som det dannes forløp på. (I F_UTD_PERSON).....	98
7.12.	Valg av identifiserende nøkkel i NUDB	98
7.12.1.	Oversikt over nøkkelalternativene.....	98
7.12.1.1.	Fødselsnummer	98
7.12.1.2.	Snr (statistikknummer).....	99
7.12.2.	Valg av løsning.....	99
7.12.2.1.	Snr-katalogen	99

7.12.2.2. BEBAS.....	99
7.12.2.3. Fil med link mellom Dnr. og Fnr.....	99
7.12.2.4. Linkmuligheter i de forskjellige kataloger/databaser:.....	100
7.12.3. Implementering av nøklene i NUDB.....	101
7.12.3.1. Benytte BEBAS/LINK-filen ved uttak.....	101
7.12.3.2. Benytte Snr-katalogen ved uttak.....	101
7.12.3.3. Løpende endring av Fnr i NUDB, og ved uttak.....	101
7.12.3.4. Benytte Snr som identifiserende id for personer i NUDB.....	102
7.12.4. Ugyldige fødselsnummer og Dnr.....	102
7.13. Studieretning 13 og 15.....	102
8. Feil og mangler i dataene. Endringer i prinsipper, definisjoner og avgrensning av massen 103	
8.1. Innledning.....	103
8.2. Feil og mangler.....	103
8.2.1. Generelt.....	103
8.2.2. 1970-1979.....	105
8.2.3. 1980-1989.....	106
8.2.4. 1990-1998.....	107
8.2.5. 1999-.....	107
8.3. Prinsipper, definisjoner og avgrensning av massen.....	109
8.3.1. Generelt.....	109
8.3.2. 1970-1979.....	110
8.3.3. 1980-1989.....	112
8.3.4. 1990-1998.....	113
8.3.5. 1999-.....	115
Vedlegg	
Prosjektskriv.....	120
Prosjektskriv fase 2.....	123
Begrepsforklaringer.....	126
En statistikkfaglig utredning av dubletter på fnr/snr i NUDB.....	132
En IT-faglig utredning av dubletter på fnr/snr i NUDB.....	134

Innledning

NUDB er et datavarehus bygd opp vha. utdanningsdata fra Seksjon for utdanningsstatistikk for årene fra og med 1970 og fram til i dag. NUDB er bygd opp etter samme modell som er benyttet av FD-Trygd i Statistisk sentralbyrå, slik at data fra NUDB forholdsvis enkelt vil kunne sammenstilles med data fra FD-Trygd. En annen mye brukt benevnelse i SSB på datavarehus av denne typen er forløpsdatabaser. I denne typen databaser vil alle hendelser som inntreffer på valgte variabler, dateres til når hendelsen fant sted. I tillegg til uttak av data for analyse av gjennomstrømninger, så muliggjør denne transformasjonen av data at en i tillegg kan ta ut tverrsnitt for ethvert tidspunkt (optimalisert for uttak per 1.okt. for hvert år) i datavarehusets gyldige tidsperiode.

NUDB inneholder per november 2007 opplysninger for hele befolkningen for perioden fra og med 1970 og fram til og med 2006, og vil bli oppdatert med nye årganger med befolknings- og utdanningsdata fortløpende i årene som kommer. Målet er at NUDB skal benyttes som den primære kilden for produksjon av fremtidig utdanningsstatistikk og for utlevering av utdanningsdata i oppdragssammenheng til forskningsformål etc. Person og personens kurs er statistisk analyseenhet for alle opplysningene i NUDB. Alle records i databasen har påkoblet SNR (kalt SNR_NUDB i NUDB) slik at det er mulig å knytte sammen informasjon fra NUDB med andre dataregistre, databaser og/eller data fra FD-Trygd. Dataene i NUDB vil være spesielt velegnet til analyse av mange aktuelle problemstillinger innen utdanningsforløp, men vil også kunne gi grunnlag for studier av overgangen mellom f.eks. utdanning og arbeid.

Styringsgruppen for NUDB har bestått av seksjonssjef for s360 Seksjon for utdanningsstatistikk, Elisabetta Vassenden, kontorsjef for s303 It-seksjonen ved avd. for personstatistikk Kristian Lønø/Tom Hansen, fra s360 Lise Styrk Hansen, Terje Risberg og Kjetil Digre.

Det daglige arbeidet med å tilrettelegge data for NUDB samt opprettelsen av forløpsdatabasen har vært utført av prosjektgruppa ved Avdeling for personstatistikk, SSB. Prosjektgruppa har bestått av personer fra både statistikkfaglig- (seksjon 360) og IT-faglig side (kontor for IT, 303). Torill Vangen på seksjon for befolknings- og utdanningsstatistikk er prosjektleder, mens Johnny Johansen fra Kontor for IT (303) har hatt det IT-tekniske ansvaret fram til mars 2002, da Anders Akselsen overtok. Prosjektgruppa har ellers bestått av: fra s303 Knut Inge Bøe, Carina Nordseth, Soheila Derakhshanfar og fra s360: Lise Styrk Hansen, Arild Rognan, Astri Halsan Høiskar og Anne Marie Holseter. I første del av prosjektet deltok også Jorun Lajord fra seksjon 350 Levekårsstatistikk.

Prosjektet har hatt både en ekstern og en intern referansegruppe. Den interne referansegruppa hadde medlemmer fra ulike seksjoner i SSB. Den eksterne referansegruppa bestod av representanter fra Utdannings- og forskningsdepartementet, Læringssenteret, Norsk institutt for studier av forskning og utdanning(NIFU) og Frisch-senteret.

Prosjektgruppen har valgt å tillegge dokumentasjonen av prosjektet stor vekt. Dette gjelder dokumentasjon av dataene som inngår i NUDB så vel som dokumentasjon av aktivitetene som har vært forbundet med å bygge opp forløpsdatabasen. I dette notatet er det dokumentert det arbeid som er utført for å spesifisere og bearbeide de ulike opplysningene om utdanningsperioder samt arbeidet med å bygge opp forløpsdatabasen NUDB etter disse spesifikasjonene. Det har vært behov for å bygge opp ett eget begrepsapparat (se vedlegg) rundt den modellen som er benyttet for å sikre lik forståelse av de problemstillingene og forslag til løsninger som har blitt skissert underveis i prosjektet.

Nedenfor følger en kort leseveiledning for resten av dette notatet:

Kapittel 1 er en introduksjon til hva NUDB er og hvordan forløpsmodellen som ligger bak fungerer. Her vil det i tillegg bli gitt en kort forklaring til de arbeidsprosesser som har inngått i NUDB.

Arbeidsprosessene som er beskrevet er allmenngyldige for alle statistikkområder som benytter FD-trygds forløpsmodell. Kapitlet kan og bør leses av alle som har tenkt å foreta uttak av data fra NUDB, samt personer som vurderer modellen for bruk i andre prosjekter under andre statistikkområder. Personer fra Kontor for IT (303) kan på forespørsel være behjelpelig med mer utdypende forklaringer ved behov.

Kapittel 2 beskriver grunnlagsregistrene som har vært benyttet som input til NUDB, samt arbeidet med å konsistenssjekke data innad på filene, samt filene seg imellom.

Kapittel 3 omhandler arbeidet med å tilrettelegge temporære arbeidsfiler som har vært benyttet som mellomsteg i arbeidet med å bygge opp NUDB. I dette notatet og ellers i all dokumentasjon som omhandler FD-Trygd-modellen er det benyttet benevnelsen 'miniregister' om disse temporære arbeidsfilene. Det gis en gjennomgang av alle variabler fra grunnlagsregistrene som tas med videre inn i NUDB, og beskrivelse av de kontroller som er blitt utført på hver enkelt variabel i denne prosessen. Videre gis det en beskrivelse av alle omkodingsprosesser som er gjennomført på de ulike kodelistene som tilhører hver enkelt variabel. Kapitlet kan leses av alle som ønsker å se på hvilke kontroller som er utført på variablene, resultatene av disse og hvordan omkodingen av eldre kodelister til nye har foregått.

Kapittel 4 inneholder diagrammer med flytskjemaer for de ulike prosessene i arbeidet med å bygge opp miniregistrene som er omtalt i kapittel 3. Kan leses som et supplement til kapittel 3.

Kapittel 5 er en faglig beskrivelse av forløpstabellene i NUDB

Kapittel 6 gir en oversikt over alle tabellene i NUDB, både forløpstabellene og andre tabeller. Alle variablene er linket mot definisjon og kodelister, som også er tilgjengelig på <http://www.ssb.no/mikrodata/>

Kapittel 7 er samling av ulike statistikkfaglige spørsmål som er gjennomgått ved opprettelsen av NUDB. Oversikten gir detaljerte statistikk- og IT-faglige vurderinger som ligger bak de valg som er foretatt mht. valg av variabler, kodelister, identifiserende nøkler etc..

Kapittel 8 gir en beskrivelse av dataene, både med hensyn til endringer i utdanningssystemet, endringer i innhentingssystemet for utdanningsdataene og en oversikt over kjente feil og mangler i dataene. Bør leses av alle som ønsker å vite noe om datagrnnlaget i NUDB.

1. Introduksjon til NUDB

1.1. Bakgrunn for modellvalg

Modellen som er benyttet i NUDB og tankegangen som ligger bak design og implementering av denne er hentet fra prosjektet FD-Trygd. I FD-trygd har det siden 1996 vært utviklet og arbeidet med en forløpsmodell som har resultert i en stor og omfattende forløpsdatabase for persondata. Eier av FD-Trygd er Seksjon for leveårsstatistikk (350). FD-trygd produserer i dag forløpsdata for en rekke temaer innenfor området trygdeytelser. I tillegg inneholder FD-Trygd mange andre typer persondata som demografi-, inntekts- og formues-, arbeidssøker- og utdanningsdata.

Når det gjelder implementasjonen av utdanningsdata i FD-Trygd så har det vært et nært samarbeid mellom seksjon 360 og prosjektgruppen for FD-Trygd. Forut for NUDB hadde det fra Seksjon 360's side vært et uttalt ønske om å samle sine utdanningsdata i en felles database som muliggjorde enkle uttak av forløps- og tverrsnittsdata for produksjon av statistikk. I tillegg skulle databasen enkelt kunne benyttes for utlevering av filer med gjennomstrømningsvariabler for forskningsformål. Erfaringene fra, - og resultatene av samarbeidet med FD-Trygd var så positive at det ble vedtatt å bygge opp NUDB etter samme modell som FD-Trygd. Vi vil i dette kapitlet forsøke å gi en kort innføring i modellen som er omtalt samt hvilke prinsipper som ligger bak den.

Arbeidet med å bygge opp et datavarehus som NUDB kan grov deles inn i tre hovedaktiviteter:

1. Tilrettelegging av grunnlagsregistrene (se kapittel 2).
2. Tilrettelegging av miniregistrene (se kapittel 3).
3. Danning av forløp i datavarehuset (se kapittel 5).

Punkt 1 og 2 er utført på fagsiden, mens punkt 3 er utført på IT-siden. Arbeidet med å bygge opp NUDB må likevel sees på som en iterativ prosess hvor fag- og IT-siden i felleskap har løst de ulike problemene som har dukket opp underveis.

1.2. Definisjon av en forløpsdatabase

1.2.1. Hva er en forløpsdatabase?

En forløpsdatabase (og et datavarehus) er egentlig ikke noe mer enn en logisk sammensetting av data som er hentet fra allerede eksisterende operasjonelle systemer (les: grunnlagsfiler for offisiell statistikk), der data er strukturert og tilrettelagt for analyse av sammenhenger mellom fortiden, nåtiden og fremtiden.

Det essensielle i en forløpsdatabase er **dateringen av hendelser**. Med hendelser menes her endringer som skjer innenfor hver enkelt variabel, samt overganger mellom viktige grupperingsvariabler (sk. hovedgrupper). En hovedgruppe kan f.eks. inneholde skolenivåene: grunnskole, videregående utdanning og høyere utdanning. I et slikt tilfelle vil vi datere overganger mellom f.eks. videregående utdanning og høyere utdanning for hvert enkelt tilfelle. I tillegg dateres alle tilganger og avganger for hvert enkelt tilfelle (som oftest person).

Nye data som tilføres/lastes inn i en forløpsdatabase, bygger på allerede "løpende" data i basen. Dvs. at nye data sammenliknes med de sist registrerte data i databasen. Regler som er bestemt på forhånd avgjør hvorvidt nye data inneholder ny og interessant informasjon for forløpsdatabasen eller ikke. Dersom det ved en slik sammenligning av datasett viser seg at det har vært endring av verdier i ett eller flere variabler så vil det genereres en **ny record** (i tillegg til allerede eksisterende records i databasen). På

denne nye recorden dateres både selve recorden som en enhet, samt de av variablene som har endret verdi. Dette er et viktig prinsipp som gjør oss i stand til å finne ut **hvilke** endringer som skjedde **når**.

1.2.2. Revisjon av forløpsdata

I mer ordinære transaksjonsdatabaser vil innholdet i databasen bli endret (eldre opplysninger blir oppdatert med nye opplysninger) etter hvert som tiden går. F.eks. innholdet i variabelen 'adresse' endre seg for en person som flytter. I forløpsdatabaser er det viktig å ta vare på de gamle verdiene (f.eks. alle tidligere registrerte adresser) på det aktuelle tidspunktet de ble registrert. Dette for å kunne "gjenskape" ethvert tidspunkt i bakover i tid.

Hva med revisjon av data i databasen da? Enkelt sagt kan en si at i tradisjonelle databaser er det lett å oppdatere/endre på eksisterende data, men noe vanskeligere å endre struktur på modellen "on the fly". For forløpsdatabaser er det motsatt; det er vanskelig/uheldig å oppdatere data i basen etter at data er lagt inn, mens det er forholdsvis enkelt å endre struktur/legge til nye variabler etc. Dette gjør at det er ekstremt viktig å legge ned mye resurser i tilpasningen og tilretteleggingen av data før de lastes inn i forløpsdatabasen.

Et eksempel: Etter at grunnlagsregister for år YYYY er lastet inn i basen, så er data fra registeret for YYYY-1 (og data for alle tidligere innlagte register) låst i databasen for "alltid". Dette siden neste kjøring (innlegging av data i forløpsdatabasen) alltid bygger videre på forrige kjøring.

1.3. Fordeler og ulemper ved den valgte modellen

Fordeler:

- Ved uttak av tverrsnittsdata for en hvilken som helst årgang benyttes ett program mot hver av maksimalt fire tabeller (KURS-, PERSON-, DEMOGRAFI- og/eller oppslagstabellen med bakgrunnsopplysninger).
- Uttak kan utføres like enkelt fra Oracle som fra SAS (PROC SQL).
- Svært mange ulike register over en lang tidsperiode er gjort tilgjengelig fra noen få tabeller i en database, hvor variablene kan tas ut med betydelig mindre innsats enn tilfellet har vært fram til i dag.
- Det er enkelt å utføre ad-hoc spørringer mot et hvilket som helst tidspunkt i datamaterialet.
- Uttak av data til tabeller/studier av gjennomstrømning i utdanningssystemet er enklere å lage.

Ulemper:

- For personer fra den statistikkfaglige siden av prosjektet kan de IT-tekniske løsningene som modellen støtter seg på, være noe vanskelig "å få tak på" til å begynne med. Det vil være en stor utfordring for deltagerne fra IT-siden i prosjektet å gjøre modellen og tankegangen bak mer forståelig for deltagere fra andre fagområder. Det er viktig å arbeide med å senke terskelen for å sikre at dialogen mellom den statistikkfaglige- og den IT-faglige siden i prosjektet foregår på et så høyt faglig plan som mulig.
- Uten et uttakssystem vil uttak av data fra et slikt datavarehus kreve dybdeforståelse av modellen for å sikre at riktige data tas ut til enhver tid. Et uttakssystem kan i sin enkleste form være et programbibliotek med ferdigskrevet kode for de mest vanlige uttakene slik som uttak av data for løpende statistikk etc.

1.4. Resultat

Denne måten å organisere data på kan ved første øyekast virke svært så plasskrevende. Det genereres tross alt nye records til databasen for hver eneste hendelse som skjer i datamaterialet! I praksis er det som regel det motsatte som er tilfelle; at datamaterialet komprimeres. Årsaken til dette er at nye data ofte

ikke gir oss **ny** informasjon (i grunnlagsregistrene er det ofte slik at nye records kun er rene gjentakelser av foregående records). I slike tilfeller registreres det ikke noen nye opplysninger i NUDB (det er ingen nye opplysninger å registrere). Samlet fører dette til at vi som regel oppnår en betydelig komprimeringsgevinst (typisk i forholdet 2:1 - 3:1) i forhold til tradisjonelle databaser og flate filer. Da grunnlagsfilene for utdanningsdata er organisert som årsfiler med betydelige endringer mellom hver årsfil er ikke komprimeringsgevinsten blitt som på en del andre statistikkområder i FD-Trygd. Bruk av modellen har likevel gitt en formidabel komprimeringsgevinst mhp. antall records: For årene 1970 og fram til i dag (når er "i dag") har det vært benyttet **ca. 115 millioner records** fra over **80 unike register** for å bygge opp datavarehuset NUDB. Ved en tilsvarende opptelling i NUDB finner vi igjen kun 75 millioner records fordelt på tre forløpstabeller og en oppslagstabell.

Resultat for NUDB

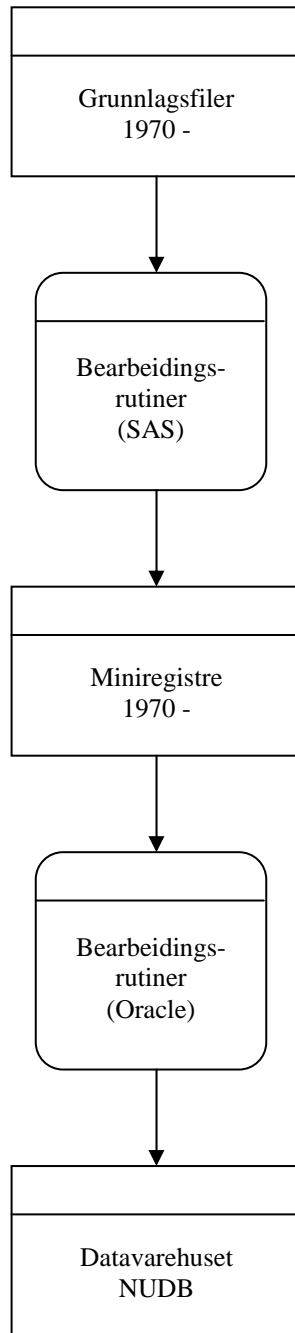
- En reduksjon på 40 millioner records fra de opprinnelige grunnlagsfilene
- Alle utdanningsdata for perioden 1970 og fram til i dag er gjort sammenlignbare og enkelt tilgjengelig for uttak av tverrsnitt og gjennomstrømning
- Det er nå mulig med ad-hoc-spørringer mot alle årganger med utdanningsdata
- Alle tverrsnitt kan gjenskapes; for hele perioden fra 1970 og fram til i dag
- Alle utdanninger er kodet med NUS2000-koder; for hele perioden fra 1970 og fram til i dag

Det er mulig (og f. hv. enkelt) å koble utdanningsdata sammens med f.eks. arbeidssøker- og sysselsettingsdata for forløpsanalyse av gjennomstrømninger mellom utdanning og arbeid.

1.5. Arbeidsflyten i NUDB prosjektet

NUDB er bygd opp av over 80 unike grunnlagsfiler (igangværende- og avsluttafiler, filer for befolkningens høyeste utdanning, demografi-filer etc.) Det har vært lagt ned et betydelig arbeid i å spesifisere og hente ut data for de variablene som nå inngår i NUDB. Grunnlagsfilene som inngår i NUDB har vært organisert som årsfiler med opplysninger knyttet til et bestemt tidspunkt. Under arbeidet med å tilrettelegge og kontrollere årsfilene slik at de blir sammenlignbare over tid, er det laget et nytt sett med filer (heretter kalt miniregistre) for de dataene som hentes ut fra de ulike årsfilene. Miniregistrene er så blitt benyttet som input til databasen hvor datatene bl.a. har vært igjennom nye konsistenskontroller i arbeidet med kvalitetssikre det arbeidet som har vært utført på miniregistrene (arbeidet med konsistenskontroller av dataene pågår kontinuerlig, også etter at dataene er lagt inn i forløpsdatabasen). Etter at dataene i miniregistrene for hver enkelt årsfil er spesifisert og kontrollert har prosjektgruppas arbeid bestått i å modellere dataene. Dette betyr kort fortalt å foreta en omorganisering av dataene, f. eks. fra data per tidspunkt, til forløpsdata. Omorganiseringen til forløpsdata krever en nøye gjennomgang av mulig og relevant datering av opplysningene i miniregistrene og i de opprinnelige årsfilene for utdanningsdata, slik at hendelser (som beskrevet over) blir ivarettatt og registrert på best mulig måte i forløpsdatabasen. I NUDB er slike hendelser registrert ved hjelp av egne datovariabler med format år, måned (YYYYMM). Etter at datamodelleringen er avsluttet har prosjektgruppas arbeid bestått i å programmere og utvikle rutiner for danning av forløp og ilegg av data i databasen.

På neste side er et diagram som viser dataflyten som er beskrevet ovenfor:



2. Tilrettelegging av grunnlagsregistrene

Det er tre hovedtyper av utdanningsopplysninger som er benyttet i NUDB: data over hvilken utdanning et individ til enhver tid er/har vært i gang med (igangværende), data over hvilken type utdanning et individ har fullført ved ulike tidspunkt (avslutta), og til slutt data over individets til enhver tids høyeste fullførte utdanning (BHU). I tillegg er det benyttet opplysninger fra diverse befolkningsfiler for å bygge opp demografiske bakgrunnsopplysninger.

I arbeidet med å transformere data fra grunnlagsregistre, -som er tverrsnittsfiler (årfiler), til forløpsdata har vi benyttet et mellomsteg hvor vi har dannet sk. miniregistre. Dette er flate filer som danner grunnlaget for hva som er lagt inn i NUDB og disse inneholder transformerte- og tilrettelagte data for de variablene som nå inngår i NUDB. I kapitlene som følger, er det en gjennomgang av hvilke spesifikasjoner som har vært lagt til grunn for utarbeidelsen av miniregistrene for de tre nevnte hovedtypene av utdanningsopplysninger. I gjennomgangen har det også vært aktuelt å omtale variable som ikke er inkludert i NUDB, men som for eksempel nyttes til kontrollformål, eller oppbygging av andre (nye) variabler.

2.1. Beskrivelse av grunnlagsregistre

Grunnlagsregistrene er årgangsfiler som danner utgangspunkt for dannelse av miniregistrene. Grunnlagsregistrene har ulik struktur og innhold. En årsak er at det stadig har tilkommet, men også forsvunnet variabler som innhentes fra utdanningsinstitusjonene. For å lette innlastingen/danning av forløp i NUDB er det laget et sett identiske filer hva angår filbeskrivelse, - selv om ikke alle årgangsfilene inneholder de aktuelle variablene. I tillegg til at årgangsfilene behandles på denne måten, gjennomgår de også ulike kontroller for å sørge for at opplysningene som lastes inn i NUDB er så korrekt som mulig. Det blir også koblet på en del nye variabler på årgangsfilene. Ut i fra allerede eksisterende variabler på årgangsfilene konstrueres det også nye variabler som skal benyttes i NUDB. Nedenfor følger en gjennomgang av grunnlagsregistrene før bearbeiding til miniregistre.

2.1.1. Filidenter

Samtlige filer er langtidslagret som flate filer på Unix, og er dokumentert i DataDok. De ulike årgangene har imidlertid noe forskjellig filbeskrivelse.

2.1.1.1. Igangværende utdanning per 1.okt.:

Årgangene 1974-1986 har samme filbeskrivelse (se DataDok for filbeskrivelse), og har følgende sti på UNIX:

Årgang	Filident UNIX
1974	\$UTD/aargang/arkiv/igang_brutto/g1974.dat
1975	\$UTD/aargang/arkiv/igang_brutto/g1975.dat
...	...
1986	\$UTD/aargang/arkiv/igang_brutto/g1975.dat

Årgangene fra 1987 og fremover har noe varierende filbeskrivelser (se DataDok for filbeskrivelser), og har følgende sti på UNIX:

Årgang	Filident UNIX
1987	\$UTD/aargang/arkiv/igang_netto/g1987.dat
1988	\$UTD/aargang/arkiv/igang_netto/g1987.dat
1989	\$UTD/aargang/arkiv/igang_netto_nus2000/g1989.dat
1990	\$UTD/aargang/arkiv/igang_netto_nus2000/g1990.dat
...	...
1999	\$UTD/aargang/arkiv/igang_netto_nus2000/g1999.dat

2.1.1.2. Avslutta utdanning i "skole"-året

Årgangene 1974-75 til og med 1985-86 har samme filbeskrivelse (se DataDok for filbeskrivelse), og ligger på følgende sti på UNIX:

Årgang	Filident UNIX
1974/75	\$UTD/aargang/arkiv/avslutta_nus73/g1974g1975.dat
1975/76	\$UTD/aargang/arkiv/avslutta_nus73/g1975g1976.dat
...	...
1985/86	\$UTD/aargang/arkiv/avslutta_nus73/g1985g1986.dat

Årgangene 1986-87 til og med 1998-99 har ulik filbeskrivelse (se DataDok for filbeskrivelser). Noen av årgangene har imidlertid samme filbeskrivelse. Filene ligger på følgende sti på UNIX:

Årgang	Filident UNIX
1986/87	\$UTD/aargang/arkiv/avslutta_nus2000/g1986g1987.dat
1987/88	\$UTD/aargang/arkiv/avslutta_nus2000/g1987g1988.dat
...	...
1998/99	\$UTD/aargang/arkiv/avslutta_nus2000/g1998g1999.dat

2.1.1.3. Befolkningens høyeste utdanning (BHU) per 1.okt.

Årgangen 1970 har egen filbeskrivelse, og ligger på følgende sti på UNIX:

Årgang	Filident UNIX
1970	\$UTD/utdnivaa/arkiv/nus2000/g1970.dat

Årgangen 1980 har også egen filbeskrivelser, og ligger på følgende sti på UNIX:

Årgang	Filident UNIX
1980	\$UTD/utdnivaa/arkiv/nus2000/g1980.dat

Årgangene 1985-1997 har lik filbeskrivelse, og ligger på følgende sti på UNIX:

Årgang	Filident UNIX
1985	\$UTD/utdnivaa/arkiv/nus2000/g1985.dat
1986	\$UTD/utdnivaa/arkiv/nus2000/g1986.dat
...	...
1997	\$UTD/utdnivaa/arkiv/nus2000/g1997.dat

Årgangen 1998 har også egen filbeskrivelser, og ligger på følgende sti på UNIX:

Årgang	Filident UNIX
1998	\$UTD/utdnivaa/arkiv/nus2000/g1998.dat

Årgangene 1999-2000 har lik filbeskrivelse, og ligger på følgende sti på UNIX:

Årgang	Filident UNIX
1999	\$UTD/utdnivaa/arkiv/nus2000/g1999.dat
2000	\$UTD/utdnivaa/arkiv/nus2000/g2000.dat

2.1.2. Filbeskrivelser for grunnlagsregistrene

I de neste kapitlene følger en gjennomgang av variabler på grunnlagsregistrene som er benyttet for oppbygging av NUDB. Gjennomgangen tar for seg variablenes navn og beskrivelse av denne. For en mer detaljert beskrivelse av variablenes posisjon på filene, lengde, datatype osv. henvises det til DataDok.

Nedenfor har vi listet opp hvilke variabler grunnlagsregistrene inneholder og hvilke variabler vi tar ut fra de ulike filene, samt forklaring på noen av variablene. Variabler som ikke inngår i miniregistrene blir ikke omtalt nærmere.

2.1.2.1. *Igangværende utdanning per 1.okt. og avslutta utdanning i løpet av året*

Igangværende filene omfatter alle personer som er i gang med en utdanning ut over ordinær grunnskole per 1.okt. (SSB har ikke individdata for igangværende utdanning på ordinær grunnskole.)

Årgangene 1974-1986 for igangværende utdanning var i utgangspunktet "bruttofiler". Det vil si at et individ kan være registrert med å være i gang med flere enn et kurs/enkeltutdanning per 1.okt.. Disse filene er gjort om til "nettofiler" i bearbeidingen til dannelse av miniregister. Det vil si at kun en av utdanningene et individ er registrert igangværende med, er tatt med videre inn i miniregisteret, og derfra videre inn i databasen. De andre igangværende-årgangene er nettofiler. Seleksjonen av hvilket kurs/enkeltutdanning som tas med videre er gjort på følgende måte:

Først velges den utdanningen som er registrert som heltid. Dersom det er flere utdanninger for et individ som er registrert som heltid, blir det kurset/enkeltutdanningen som hadde det høyeste klassetrinnet valgt ut, og tatt med videre til databasen. Dersom det ennå er flere kurs/enkeltutdanninger registrert som igangværende for samme individ, blir det kurset som har det høyeste utdanningsnivået av disse valgt ut, og tatt med videre inn i basen. Utdanningsnivået bestemmes av første siffer i NUS73. NUS73 er utdanningsstandarden som ble benyttet på disse årgangene. Dess høyere siffer, dess høyere utdanningsnivå. Denne utvelgelsesstrategien viser seg å fjerne samtlige "dubletter", slik at vi sitter igjen med bare et igangværende kurs/enkeltutdanning per individ.

Filårgangene fra 1987 til nyeste fil er nettofiler, og utvelgingen av hvilken utdanning som skal være med er allerede foretatt før tilretteleggingen av filene til miniregistrene.

Avsluttafilene omfatter alle personer som avslutter grunnskolen, eller avslutter/avbryter videregående skole eller høyere utdanning. Avsluttafilene for år t inneholder opplysninger om avsluttet utdanning i perioden 1.okt. år t-1 til 30.sept. år t. Vi starter med opplysningene fra avsluttafil 1974/75.

Avsluttafilene er bruttofiler. Det vil si at alle kurs/enkeltutdanninger som en person avslutter (fullfører eller avbryter) i løpet av et år blir registrert i fila.

Nedenfor følger en gjennomgang av hvilke variabler som finnes på årgangsfilene:

Utdanningens art

Denne variabelen tilsvarer variabelen utdanningens art (omkodet) i filen for høyeste fullførte utdanning. Kodene følger "Standard for utdanningsgruppering (NUS)". NUS finnes i flere utgaver, og det varierer hvilken utgave som ligger på filene. Disse har NUS73. På grunn av dette må filene gjennom en omkodingsprosess, da vi kun skal ha med kodene fra de to siste utdanningsstandardene NUS89 og NUS2000. NUS73 kodes først om til NUS89, og deretter fra NUS89 til NUS2000

Klassetrinn

For studenter ved universitetene har fagseksjonen et eget program for oppjustering av klassetrinn. For videregående skoler og høyskoler settes klassetrinn av skolen med utgangspunkt i kurset eleven/studenten går på. Hvis studenten skifter mellom høyskole og universitet, eller endrer studiemål, kan klassetrinn derimot være mindre representativ for den faktiske progresjonen studenten har. I

NUS2000 er klassetrinnene annerledes enn i tidligere versjoner av NUS. Miniregistrene vil derfor pålegges disse klassetrinnene om de ikke skulle finnes på grunnlagsregistrene. Filene 1974-1986 har klassetrinn som korresponderer med NUS73. Variabelen kodes derfor om i bearbeidingen. Dette gjøres ved å først kode om NUS73 klassetrinnene til NUS89 klassetrinn. Deretter kodes disse om til NUS2000 klassetrinn.

Heltid/deltid

Dette er en variabel som i utgangspunktet blir satt av skolen ut i fra om kurset er ment å utgjøre et fulltidsstudium. På Avsluttafilene blir alle studenter ved universitetet kodet som heltidsstudenter. På igangværendefilene blir imidlertid universitetsstudentene registrert på heltid/deltid etter gitte kriterier - f.eks. forventet studieprogresjon.

Kursets påbegynnningstidspunkt (Mnd/år)

Tas med dersom det ikke finnes opplysninger om starttidspunkt på kjennemerket 'Elevens påbegynnningstidspunkt'

Kursets avslutningstidspunkt (Mnd./år)

Tas med dersom det ikke finnes opplysninger om sluttidspunkt på kjennemerket 'Elevens avslutningstidspunkt'

SSBs skolenummer

Denne variabelen endrer navn til 'inr' i miniregistrene.

Klassebetegnelse

Ikke i NUDB

Voksenopplæring

Variabelen viser om kurset er en utdanning som går inn under lov om voksenopplæring. Variabelen er ikke med i dataene fra 2000 Variabelen har kodene:

Voksenopplæring = 1

Ikke voksenopplæring = 2

Hjelpenummer for RVO-kode

ikke i NUDB

Linjenummer

ikke i NUDB

Fødselsnummer

Elevens-/studentens fødselsnummer.

Utfall

Variabelen viser om personen har avbrutt eller fullført kurset. Hovedregelen er at utdanning regnes som fullført når vitnemål utstedes. For noen kurs blir det imidlertid ikke utstedt vitnemål. Disse blir også registrert som fullført. Kodene for variabelen er:

Ikke fullført/Avbrutt = 2

Fullført = 8

På igangværendefilene 1974-1986 settes utfall til blank.

Utfall detaljert

Variabelen er mer detaljert enn Utfall og varierer mellom ulike skoleslag. Variabelen er blank før 1995, og bare delvis utfyllt senere. Kodene for variabelen er:

A = Annen vurdering. Har fullført kurs særskilt tilrettelagt innenfor læreplanverket. Fullført individuelt opplegg. Utfall=2. Kilde 20

B = fullført med vitnemål/årskursbevis, bestått. Utfall 8. Kilde20

C = Ikke sensurert. Utfall=2. Kilde 40,41

D = Ikke bestått. Utfall=2. Kilde 40,41
F = Bestått fagprøve. Utfall = 8. Kilde 21
G = Meget bra bestått fagprøve. Utfall = 8. Kilde 21
I = Fullført med vitnemål/årskursbevis, ikke bestått. Utfall =8 i Reform94, utfall=2 tidligere. Kilde20. (i ettertid endret i NUDB til utfall 2, avbrutt, også for R94)
M = Mangler vitnemål/årskursbevis. Har fullført uten vurdering i ett eller flere fag, f.eks. ved stort fravær. Utfall=2. Kilde20
N = Ikke bestått fagprøve. Utfall = 2. Kilde 21
O = Alternativ VKII i skole. Utfall = 2. Kilde 20
P = Bestått: Utfall=8. Kilde 40,41
S = Sluttet/avbrutt i løpet av året. Utfall2. Kilde20
X = Ikke utfylt fra skole. Utfall = 8. Kilde 20
2 = Ikke fullført. Utfall=2. Kilde 10,21,22,23,24 og 42-46
8 = Fullført. Utfall=8. Kilde 10,21,22,23,24 og 42-47

Elevens påbegynningsstidspunkt mmåå

Gir opplysninger om når eleven påbegynte gjeldende utdanning

Elevens avslutningsstidspunkt mmåå

Gir opplysninger om når eleven avsluttet gjeldende utdanning

Skolekommune

Gir opplysninger om hvilken kommune skolen befinner seg i

Næringskode

Denne variabelen endrer navn i miniregistrene. For årganger fra 1974-1977 kalles den for 'SN72'. For årganger fra 1978-1993 kalles den 'SN78', og fra årgangen 1994 og fremover kalles den 'SN94'.

Skolegruppe har samme innhold som i variabelen i avsnittet over

Denne variabelen endrer også navn i miniregistrene. For årganger fra 1974-1977 kalles den for 'SN72'. For årganger fra 1978-1993 kalles den 'SN78', og fra årgangen 1994 og fremover kalles den 'SN94'.

Eierforhold, uomkodet

Bostedskommune

Gir opplysninger om i hvilken kommune eleven bor i. Dette kan være forskjellig fra skolekommune.

Kjønn

Mann = 1
Kvinne = 2

Alder

Alder1, omkodet

Alder2, omkodet

Påbegynnelsestidspunkt

Danner kjennemerket 'startd' for årgangene 1974-1986

Avslutningsstidspunkt

Danner kjennemerket 'sluttd' for årgangene 1974 – 1986

Elevstatus

Bosted kontra skolekommune

Studieretningskode

Variabelen benyttes til å lage nye variabler på miniregistrene ('Studretn', 'Kurstrin'). 'Studretn' lages av de to første sifrene i koden, mens 'Kurstrin' lages av det femte sifferet.

Utdanningens art, ny inndeling katalog

Variabelen benyttes til å lage nye variabler på miniregistrene ('Studretn', 'Kurstrin'). 'Studretn' lages av de to første sifrene i koden, mens 'Kurstrin' lages av det femte sifferet.

Eierforhold, omkodet

Gir opplysninger om skolen er statlig, fylkeskommunal/kommunal eller privat:

Skolekommunetype

Bostedskommunetype

Elever, påbegynt, omkodet

Elever, avsluttet, omkodet

Omkodet skolegruppe

Skoleslag, omkodet

Gir opplysninger om hvilken skole det her er tale om; grunnskole, videregående-skole, eller universitet/høgskole.

Grunnskole = 1
Videregående-skole = 2
Universitet/høgskole = 3

ISCED

Dette er den gamle ISCED-koden (1976)

Fødeland

Statsborgerskap

Fordyp

Fordypningsfag

NUS2000

NUS2000-koden

Kltrinn2000

Klassetrinn etter NUS2000-koden

Kodetype

Gir opplysninger om dette er en samlekode eller en enkeltutdanning i NUS2000

Samlekode = 1
Enkeltutdanning = 2

Uhg2000

Gruppering av universitets- og høgskoleutdanning. Variabelen skifter navn til 'Uhgruppe' i miniregisteret
Det ser ut som noen variable er listet opp 2 ganger - se ISCED på forrige side

I97ISCED

Den internasjonale utdanningskoden ISCED97

I97destn

ISCED97, tilleggsdimensjonen 'Destination'

I97orien

ISCED97, tilleggsdimensjonen 'Orientation'

I97varig

ISCED97, tilleggsdimensjonen 'Varighet'

I97grads

ISCED97, tilleggsdimensjonen 'Gradsstruktur'

Naering

Næringskode 1994

Høgskole

Gir opplysninger om hvilket høgskolesentra utdanningen er tatt ved

Innvkat**Landbak****Hskode**

Gir opplysninger om hvilken høgskole studiet er tatt på

Kompetanse i videregående

Gir opplysninger om hvilken kompetanse eleven har fått etter sin utdanning. Gjelder videregående utdanning og kan kun benyttes for data fra og med 2000.

Utd

Gruppering av skoleslag. Gir opplysninger om hvilken type skole utdanningen er tatt ved.

Studretn

Studieretning. Gir opplysninger om hvilken studieretning på videregående skole gjeldende utdanning er tatt ved

Kurstrin

Kurstrinnskode. Gir opplysninger om hvilket kurstrinn utdanningen er tatt på

Et av formålene med NUDB er å finne den perioden en person er under utdanning. Vi har derfor tatt med oss både elevens/studentens og kursets start- og sluttdato, men har bare benyttet kursets start- og sluttdato dersom elevens/studentens datoer ligger med blank. Det vil si at det bare er én start, og én sluttdato som er tatt med videre. På igangværendefilene har vi kun benyttet startdatoen fra filene. Vi har kun vært interessert i å finne ut når eleven/studenten påbegynte utdanningen sin på disse filene.

Variabelen 'Skolegruppe' eller 'Næringskode' angir "skoletype" og tar utgangspunkt i skolen, ikke hvilke kurs studentene leser. Variabelen har to sifre, og dette tilsvarer de to siste sifrene i den femsifrede koden i Standard for næringsgruppering (for undervisningssektoren) som gjaldt til og med 1992. Denne

variabelen skal benyttes til å finne ut hvilke studenter som studerer ved universitetene. Dette har betydning for å kode hovedgruppevariabelen. Vi lager også en variabel på miniregistrerne, kalt 'Univ' som markerer for de som studerer på universitetene.

2.1.2.2. Befolkningens høyeste utdanning per 1.oktober

Registeret over befolkningens høyeste utdanning omfatter personer registrert bosatt i Norge per 1.okt., som per 31/12 er 16 år og over. I tillegg registreres 15-åringer som har fullført grunnskolen eller som er i gang med en utdanning utover grunnskolenivå.

Opplysningene om høyeste fullførte utdanning blir oppdatert hvert år med opplysninger fra individ-filene over avsluttet utdanning. Fra og med filen per 1.okt. 1993 blir filen også oppdatert med opplysninger om utdanning fullført i utlandet (opplysninger fra Statens lånekasse for utdanning). Opplysningene om utenlands-utdanning dekker utdanning fullført i utlandet, av lånekassens kunder, fra og med skoleåret 86/87. Våren 1991 ble det hentet inn på skjema opplysninger om eksamen avlagt i utlandet av utenlandsfødte personer som hadde innvandret til Norge for første gang mellom 1/11 1980 og 31/12 1990, og som var 16 år og over ved utgangen av første innvandringsåret. Denne undersøkelsen het "Utdanning fullført i utlandet", og en tilsvarende undersøkelse ble også gjennomført i 1999. Respondentene i den siste undersøkelsen var alle med utenlandsk bakgrunn som var registrert bosatt i Norge og som sto med uoppgitt utdanning i registeret over befolkningens høyeste utdanning. Også frafallsgruppen i undersøkelsen fra 1991 var inkludert. I tillegg ble nordmenn som bodde i utlandet da BHU ble opprettet i 1970 som sto med uoppgitt BHU tatt med i undersøkelsen i 1999.

Det har ikke vært mulig å utarbeide BHU-status for 1981, 82, 83 og 84. Dette skyldes tekniske problemer knyttet til omleggingen av Det sentrale personregisteret. SSB har derimot utarbeidet BHU-status hvert år siden 1985, basert på bearbeiding av filene for avsluttet utdanning. Dagens register tillater en løpende ajourføring.

BHU-filene er nettofiler i det personer er enhet. Personer som har fullført mer enn en utdanning, er bare registrert med den høyeste av utdanningene. Det vil si at den utdanningen som har det høyeste nivået, er registrert. Hvis en person har fullført flere utdanninger på samme nivå, er det den utdanningen med høyeste fullføringstidspunkt som er registrert.

Alle utdanninger i SSBs utdanningsstatistikk har fått sekscifrede utdanningskoder etter den norske standarden for utdanningsgruppering (NUS). Dette gjelder også BHU. Utdanningsstandard ble første gang utarbeidet i 1970, og deretter revidert i 1973, 1989 og nå sist i 2000 (NUS2000). Ved publisering av SSBs statistikk over BHU benyttes det to alternative utdanningsvariabler for å beskrive fullført utdanning. Dette har sin bakgrunn i overgangen fra sjuårig folkeskole til niårig grunnskole. I den uomkodede utdanningsvariabelen er grunnleggende ettårig fagutdanning avlagt før 1972-73 basert på sjuårig folkeskole og ett års framhaldsskole plassert på grunnskolenivå. I den omkodede utdanningsvariabelen er tilsvarende utdanning plassert på videregående skolenivå. Hvis man vil belyse utdanningsnivå etter utdanningens kompetanse, bør den omkodede variablene benyttes. Vi har valgt å kun legge inn omkodet utdanningsvariabler inn i Nasjonal utdanningsdatabase (NUDB).

I tillegg til NUS registreres også den internasjonale standarden for utdanningsgruppering International Standard Classification of Education (ISCED), som ble utarbeidet av UNESCO på 1970-tallet og ble sist revidert i 1997 (ISCED97). I forbindelse med revisjonen av den norske standarden for utdanningsgruppering er det utarbeidet en nøkkel mellom den og ISCED97, der hver NUS-kode har en henvisning til en internasjonal kode med tilhørende tilleggsdimensjoner.

Et kryss indikerer at variabelen tas med inn i miniregistrerne for BHU.

Filbeskrivelse av BHU-1970 og hvilke variabler som tas med inn i miniregisteret

BHU-70	Variabelnavn	Variabelbeskrivelse
x	fnr	Fødselsnummer
x	kommnr	Bostedskommune
	kjoenn	Kjønn. 1 = mann, 2 = kvinne
	NUS73u	NUS73 uomkodet
	kltr73	Klassetrinn73
x	nusomkod	NUS89 omkodet. Skifter navn til 'NUS89' i basen
	kltrinno	Klassetrinn89 omkodet
x	NUS2000_o	NUS2000 omkodet. Skifter navn til 'NUS2000' i basen
x	kltrinn2000_o	Klassetrinn2000 omkodet. Skifter navn til 'kltrinn2000' i basen
x	uhgruppe_o	Universitets og høgskolegruppering, skifter navn til 'uhgruppe' i NUDB
x	I97ISCED	ISCED97-kode, omkodet
x	I97destn	ISCED97, tilleggesdimensjonen destination, omkodet
x	I97orien	ISCED97, tilleggesdimensjonen orientation, omkodet
x	I97varig	ISCED97, tilleggesdimensjonen varighet, omkodet
x	I97grads	ISCED97, tilleggesdimensjonen gradsstruktur, omkodet

Filbeskrivelse av BHU-1980 og hvilke variabler som tas med inn i miniregisteret

BHU-80	Variabelnavn	Variabelbeskrivelse
x	fnr	Fødselsnummer
x	kommnr	Bostedskommune
	flyttat	Dato for flytting
	alderu	Alder ved utgangen av året
	kjoenn	Kjønn. 1 = mann, 2 = kvinne
	opphold	Dato for første opphold i Norge
x	statborg	Statsborgerskap. NB er blank i denne årgangen
	fodeland	Fødeland
x	nusukod	NUS89 uomkodet
	kltrinno	Klassetrinn NUS89 uomkodet
x	nusomkod	NUS89 omkodet. Skifter navn til 'NUS89' i basen
	kltrinno	Klassetrinn89 omkodet
	fulland	Fullføringsland
	fulldato	Fullføringstidspunkt (YYYYMM)
	isced	ISCED-kode
x	pedsem	Fullført praktisk-pedagogisk utdanning
	nusigang	Utdanningens art NUS89, igangværende utdanning
	kltrinn	Klassetrinn igangværende utdanning, NUS89
	heldel	Heltid/deltid. 1 = heltid, 2 = deltid
	skolekom	Skolekommune
	start	Påbegynnningstidspunkt
x	NUS2000_o	NUS2000 omkodet. Skifter navn til 'NUS2000' i basen
	NUS2000u	NUS2000 uomkodet
x	kltrinn2000_o	Klassetrinn2000 omkodet. Skifter navn til 'kltrinn2000' i basen
	nus2000i	NUS2000, igangværende utdanning
	kltrinn2000i	Klassetrinn NUS2000 igangværende utdanning
x	uhgruppe_o	Universitets og høgskolegruppering, omkodet. Skifter navn til 'uhgruppe' i NUDB
	uhgruppe_i	Universitets og høgskolegruppering, igangværende utdanning
x	I97isced	ISCED97-kode, omkodet

BHU-80	Variabelnavn	Variabelbeskrivelse
x	I97destn	ISCED97, destination, omkodet
x	I97orien	ISCED97, orientation, omkodet
x	I97varig	ISCED97, varighet, omkodet
x	I97grads	ISCED97, gradsstruktur, omkodet

Filbeskrivelse av BHU-1985-97 og hvilke variabler som tas med inn i miniregisteret

BHU-85-97	Variabelnavn	Variabelbeskrivelse
x	fnr	Fødselsnummer
x	kommnr	Bostedskommune
	flyttat	Dato for flytting
	alderu	Alder ved utgangen av året
	kjonn	Kjønn. 1 = mann, 2 = kvinne
	opphold	Dato for første opphold i Norge
x	statborg	Statsborgerskap. Er blank i 1985, og er koblet på årgangene fra 1986 til 1989. Opprinnelig på filene fra 1990 til 1998.
	fodelad	Fødeland
	nusukod	NUS89 uomkodet
	kltrinnu	Klassetrinn NUS89 uomkodet
x	nusomkod	NUS89 omkodet. Skifter navn til 'NUS89' i basen
	kltrinno	Klassetrinn89 omkodet
	fulland	Fullføringsland
	fulldato	Fullføringstidspunkt
	isced	ISCED-kode, knyttet til 'nusomkod'
x	pedsem	Fullført praktisk-pedagogisk utdanning
	nusigang	Utdanningens art NUS89, igangværende utdanning
	kltrinn	Klassetrinn igangværende utdanning, NUS89
	heldel	Heltid/deltid. 1 = heltid, 2 = deltid
	skolekom	Skolekommune
	start	Påbegynnningstidspunkt
x	NUS2000_o	NUS2000 omkodet. Skifter navn til 'NUS2000' i basen
	NUS2000u	NUS2000 uomkodet
x	kltrinn2000_o	Klassetrinn2000 omkodet. Skifter navn til 'kltrinn2000' i basen
	nus2000i	NUS2000, igangværende utdanning
	kltrinn2000i	Klassetrinn NUS2000 igangværende utdanning
x	uhgruppe_o	Universitets og høgskolegruppering, omkodet. Skifter navn til 'uhgruppe' i NUDB
	uhgruppe_i	Universitets og høgskolegruppering, igangværende utdanning
x	I97isced	ISCED97-kode, omkodet
x	I97destn	ISCED97, destination, omkodet
x	I97orien	ISCED97, orientation, omkodet
x	I97varig	ISCED97, varighet, omkodet
x	I97grads	ISCED97, gradsstruktur, omkodet

Filbeskrivelse av BHU-1998 og hvilke variabler som tas med inn i miniregisteret

BHU-98	Variabelnavn	Tekst
x	fnr	Fødselsnummer
x	kommnr	Bostedskommune
	flyttat	Dato for flytting
	alderu	Alder ved utgangen av året

BHU-98	Variabelnavn	Tekst
	kjonn	Kjønn. 1 = mann, 2 = kvinne
	opphold	Dato for første opphold i Norge
x	statborg	Statsborgerskap. Er blank i 1985, og er koblet på årgangene fra 1986 til 1989. Opprinnelig på filene fra 1990 til 1998.
	fodelad	Fødeland
	nusukod	NUS89 uomkodet
	kltrinnu	Klassetrinn NUS89 uomkodet
x	nusomkod	NUS89 omkodet. Skifter navn til 'NUS89' i basen
	kltrinno	Klassetrinn89 omkodet
	fulland	Fullføringsland
	fulldato	Fullføringstidspunkt
	isced	ISCED-kode, knyttet til 'nusomkod'
x	pedsem	Fullført praktisk-pedagogisk utdanning
	nusigang	Utdanningens art NUS89, igangværende utdanning
	kltrinn	Klassetrinn igangværende utdanning, NUS89
	heldel	Heltid/deltid. 1 = heltid, 2 = deltid
	skolekom	Skolekommune
	start	Påbegynnningstidspunkt
x	NUS2000_o	NUS2000 omkodet. Skifter navn til 'NUS2000' i basen
	NUS2000u	NUS2000 uomkodet
x	kltrinn2000_o	Klassetrinn2000 omkodet. Skifter navn til 'kltrinn2000' i basen
	nus2000i	NUS2000, igangværende utdanning
	kltrinn2000i	Klassetrinn NUS2000 igangværende utdanning
x	uhgruppe_o	Universitets og høgskolegruppering, omkodet. Skifter navn til 'uhgruppe' i NUDB
	uhgruppe_i	Universitets og høgskolegruppering, igangværende utdanning
x	I97isced	ISCED97-kode, omkodet
x	I97destn	ISCED97, destination, omkodet
x	I97orien	ISCED97, orientation, omkodet
x	I97grads	ISCED97, gradsstruktur, omkodet
x	I97varig	ISCED97, varighet, omkodet

Filbeskrivelse av BHU-1999 og hvilke variabler som tas med inn i miniregisteret

BHU99	Variabelnavn	Variabelbeskrivelse
x	fnr	Fødselsnummer
x	kommnr	Bostedskommune
	alder	Alder ved utgangen av året
	kjoenn	Kjønn. 1 = mann, 2 = kvinne
	fodato	Dato for første opphold i Norge
x	statborg	Statsborgerskap
	fodeland	Fødeland
	invkat	Innvandringskategori
	landbak	Landbakgrunn
	NUS2000u	NUS2000 uomkodet
	kltrinn2000u	Klassetrinn NUS2000 uomkodet
x	NUS2000_o	NUS2000 omkodet. Skifter navn til 'NUS2000' i basen
x	kltrinn2000_o	Klassetrinn2000 omkodet. Skifter navn til 'kltrinn2000' i basen
x	fulland	Fullføringsland
x	fulldato	Fullføringstidspunkt (YYYYMM)
x	pedsem	Fullført praktisk-pedagogisk utdanning

BHU99	Variabelnavn	Variabelbeskrivelse
x	kildebhu	Kilde for BHU-opplysning
x	I97isced	ISCED97-kode, omkodet
x	I97destn	ISCED97, destination, omkodet
x	I97orien	ISCED97, orientation, omkodet
x	I97varig	ISCED97, varighet, omkodet
x	I97grads	ISCED97, gradsstruktur, omkodet
x	uhgruppe	Universitets og høyskolegruppering
	NUS2000_i	Utdanningens art NUS2000, igangværende utdanning
	kltrtinn_i	Klassetrinn igangværende utdanning, NUS2000
	heldel_i	Heltid/deltid, igangværende utdanning
	orgnr_i	Organisasjonsnummer, igangværende utdanning
	skolekom_i	Skolekommune, igangværende utdanning

Ikke alle BHU-filer har alle variablene det er krysset av for. Imidlertid får de "koblet på" variabelen i miniregisteret med opplysninger der vi har kunnet framskaffe dem, og er blank der opplysningene mangler for hele årgangen.

Fra og med årgang 2000 blir BHU-filene ikke lastet inn i basen, fordi fra og med dette året vil alle data som er benyttet til å oppdatere Bhu også ligge på årgangsfilen for avsluttet utdanning.

2.2. Korreksjoner og kontroller

2.2.1. Fødselsnummerkontroll

Det er kjørt samme type fødselsnummerkontroll for hele NUDB. Programmet for dette ligger på: \$UTD/nudb/prog/sas/mrk_fnr.sas. Siden avsluttafila inneholder dubletter mht. FNR, så kjøres fødselsnummerkontrollen for disse data på en litt annen måte (se neste punkt). Programmet fjerner dubletter kun på grunnlag av fødselsnummer og beholder den første recorden av de like fødselsnummerene. I tillegg markerer programmet ugyldige fødselsnummer med variabelen 'mrk_fnr'.

Kodelisten for mrk_fnr ser slik ut:

0 = Gyldig fødselsnummer

1 = Gyldig d-nummer

2 = Ugyldig fødselsnummer som består av blankt personnummer

3 = Ugyldig fødselsnummer som ikke omfattes av 1 og 2

2.2.2. Dublettkontroller

Utføres på filene for å luke ut uønskede dubletter. Nedenfor er beskrivelser av de algoritmer som er benyttet på de forskjellige filene for å løse problemene med uønskede dubletter.

2.2.2.1. Igangværendefilene:

På igangværendefilene skal det ikke forekomme dubletter i fødselsnummer. Dublettene her defineres derfor bare ved fødselsnummer alene, og fjernes derfor i fødselsnummerkontrollen.

2.2.2.2. Avsluttafilene:

Registrene for avslutta utdanning er bruttofiler, og har derfor fødselsnummerdubletter. Det skal ikke forekomme to like kurs (lik NUS-kode) som løper samtidig i NUDB. Like kurs som er overlappende i tid slås sammen. Det vil si at vi tar startdato fra det kurset som har startet først, og sluttdato fra det kurset som er avsluttet sist. Sorteringen som er benyttet til å kontrollere records mot hverandre er:

fødselsnummer x utdanningens art x startdato x sluttdato (FNR x NUS89 x STARTD x SLUTTD). Dette kan f.eks. se slik ut, med tre records, hvor start- og endepunktene på linjene angir start- og sluttdato:

```
Record 1:      |-----|
Record 2:           |-----|
Record 3:           |-----|
```

Vi tar startdatoen fra den første recorden, og sluttdatoen fra record 3. Internvariablene hentes fra den recorden som har utfall = 8 (fullført). Det betyr at hvis kun record 2 har utfall = 8, så hentes internvariablene fra denne recorden. Hvis både record 2 og record 3 har utfall = 8, så hentes internvariablene fra record 3.

Vi merker hvilke record vi henter opplysninger fra ved slik sammenkobling i **Mrk_dl3** etter følgende regler:

Hvis start- og sluttdato (dvs. alle variabler) er hentet fra samme record, **så Mrk_dl3 = 0.**

Hvis startdato, men ikke sluttdato, hentes fra den recorden hvor de andre opplysningene er hentet fra, **så Mrk_dl3 = 1.**

Hvis stoppdato, men ikke startdato, hentes fra den recorden hvor de andre opplysningene er hentet fra, **så Mrk_dl3 = 2.**

Hvis start- og stoppdato hentes fra en record mens de andre opplysningene hentes fra en annen record, **så Mrk_dl3 = 3.**

Hvis startdato, stoppdato og de andre opplysningene hentes fra tre ulike records, **så Mrk_dl3 = 4.**

Ved sammenslåing av kurs til en periode blir Mrk_dl1 og Mrk_dl2 oppdatert med riktige koder i forhold til den nye perioden.

2.2.2.3. BHU-filene:

På BHU-filene skal det ikke forekomme dubletter i fødselsnummer. En person skal kun være registrert med den høyeste fullførte utdanningen. Har en person avlagt to eksamener på sammen nivå skal den med det høyeste fullføringstidspunktet velges. Det viste seg imidlertid at en av filene inneholdt dubletter. Dublettkontroll utføres derfor før fødselsnummerkontrollen. Denne kontrollen sletter dubletter på grunnlag av laveste NUS-kode og deretter etter laveste klassetrinn. Dublettkontrollen som er innbakt i fødselsnummerkontrollen utføres med fødselsnummer som eneste variabel, og er derfor ikke nøyaktig nok.

2.2.3. Datokontroll

Datoene som ligger på registrene er månedsopplysninger, dermed får vi ingen "ugyldige" datoer, men vi sjekker gyldige kombinasjoner av år og måned. Datoer i utdanningsfilene er på formen MMYYY. Disse er kodet om til formatet: YYYYMM.

2.2.3.1. Igangværende- og avsluttafilene:

Vi sitter da igjen med bare én start- og sluttdato per record for igangværende og avslutta filene. Vi velger på følgende måte:

Hvis elevens/studentens påbegynnings- (avslutnings-) tidspunkt er utfyllt **Så** start- (slutt-) tidspunkt settes lik elevens/studentens påbegynnings- (avslutnings-) tidspunkt **Ellers** start- (slutt-) tidspunkt settes lik kursets påbegynnings- (avslutnings-) tidspunkt.

Vi legger så inn en kode som viser hvilke datoer vi har benyttet. På igangværendefilene får vi da følgende koder for variabelen **Mrk_dl**:

- 1= startdato er hentet fra elevens/studentens påbegynningsstidspunkt
- 3= startdato er hentet fra kurssets påbegynningsstidspunkt
- 5= startdato er satt ut fra det året fila er gyldig for, da både 'Estart' og 'Kstart' er blanke. Universitetsstudentene får start i september, mens høyskolestudentene får start i august.
- 6= startdato er satt til året fila er gyldig for (md. er 08 el. 09 avhengig av om det er elever/høyskolestudenter eller universitetsstudenter) da den er satt til uoppgitt eller har "feil" dato (utenom filas gyldighetsområde). Det vil si at elevens/studentens eller kurssets påbegynningsstidspunkt er "feil". I de fleste tilfeller er det her tale om 'Estart' eller 'Kstart' = '190001' eller '190101'. Dette er universitetsstudenter.

På avsluttafilene får vi følgende koder for variabelen **Mrk_dl1**:

- 1 = Elevens/studentens påbegynnings- og avslutningstidspunkt er benyttet
- 2 = Elevens/studentens påbegynnings- og kurssets avslutningstidspunkt er benyttet
- 3 = Kurssets påbegynnings- og elevens/studentens avslutningstidspunkt er benyttet
- 4 = Kurssets påbegynnings- og avslutningstidspunkt er benyttet
- 5 = Påbegynningsstidspunktet mangler (helt), mens elevens/studentens avslutningstidspunkt er benyttet
- 6 = Påbegynningsstidspunktet mangler (helt), mens kurssets avslutningstidspunkt er benyttet
- 7 = Påbegynningsstidspunktet er uoppgitt, mens elevens/studentens avslutningstidspunkt er benyttet
- 8 = Påbegynningsstidspunktet er uoppgitt, mens kurssets avslutningstidspunkt er benyttet
- 9 = Påbegynningsstidspunktet er større eller lik sluttdato, mens elevens/studentens avslutningstidspunkt er benyttet
- A = Påbegynningsstidspunkt er større eller lik sluttdato, mens kurssets avslutningstidspunkt er benyttet

I tillegg er startdatoer som '190001' eller '190000' kodet om til det året fila er gyldig for, og gitt startmåned i august eller september, avhengig av om det er høyskole- eller universitetsstudenter det er tale om. Disse er imidlertid ikke merket annet enn med den Mrk_dl1 koden de fikk når start- og sluttdatoen ble satt i utgangspunktet. For årgangene 1974/75 t.o.m. 1985/86 har vi benyttet variablene 'paabeg' og 'avslut' for å sette 'startd' og 'sluttd'. Disse kjennemerkene viser seg å være av bedre kvalitet enn 'estart', 'kstart', 'eslutt' og 'kslutt'. Disse får ingen merking da alle datoer er registrert.

2.2.3.2. BHU-filene:

På BHU-filene ligger det opplysning om når utdanningen er fullført, med formatet YYYYMM. Disse blir kontrollert for ugyldige records, uoppgitt fullførtdato eller uoppgitt månedsopplysning (de som har blank eller 00 i månedsopplysning). Alle endringer som gjøres blir merket etter følgende regler:

Datoene som ligger på utdanningsregistre er år og månedsopplysninger, dermed får vi ingen "ugyldige" datoer, men vi sjekker gyldige kombinasjoner av år og måned.

Kontrollen som utføres på variabelen **fullføringstidspunkt** ('fulldato'):

Hvis fullføringstidspunkt ikke endres **Så**

settes mrk_dl_bhu lik 0

Hvis manglende fullføringstidspunkt **Så**

settes datoen til filens år og måned 06 (YYYY06), og mrk_dl_bhu settes lik 1

Hvis månedsopplysningen er '00' eller blank **Og** år er gyldig **Så**

settes månedsopplysning til '06' (YYYY06) og mrk_dl_bhu settes lik 2

Hvis fullføringstidspunktet er korrigert for ugyldige datoer (f. eks etter filas årgang) **Så**

settes fullføringstidspunktet til filas årgang og '09', og mrk_dl_bhu settes lik 3

Hvis fullføringstidspunktet mangler for hele fila (gjelder kun for BHU-1970) **Så**

settes fullføringstidspunktet til 197011, og mrk_dl_bhu settes lik 4.

Med andre ord så er ingen record med ugyldige datoer slettet fra fila, men endret til gyldige datoer og merkes med variabelen *mrk_dl_bhu*.

BHU-tall blir publisert som om det var fra 1.okt. hvert år, derfor korrigeres filene slik at den høyeste dato på filene blir YYYY09. Med andre ord det høyeste fullføringstidspunktet for eksempel BHU-1985 blir da 198509. BHU-tall for 1970 som er hentet fra Folke- og bolig tellingen i 1970 (FoB70) har imidlertid 197011 som avslutningstidspunkt, da dette var tidspunktet for innhenting av opplysninger ved FoB70. Fullføringstidspunktet 198011 som var på BHU-1980 opprinnelig må ha blitt satt i forbindelse med FoB80, da det ikke var noen rubrikk på spørreskjemaet som spurte om fullføringstidspunkt og ble dermed alle satt til 198011. Vi har imidlertid valgt å endre alle som hadde fullføringstidspunkt i 198011 på BHU-1980 til 198009. Tilganger mellom 1970 og 1980 kommer ikke med før i 1980 da vi ikke har BHU-filer for årene 1971-1979. Publiserte tall for BHU-status hvert år inneholder ikke personer under 16 år, men 15-åringer som har fullført grunnskolen eller som er i gang med en utdanning utover grunnskolenivå ligger registrert på BHU-filene. BHU1980 inneholder ingen 15 åringer. For å se om de blir inkludert ved neste BHU-status som var i 1985 undersøkte vi om det fantes noen på BHU-1985 filen som var 15 år i 1980. Det var en rekke personer som falt inn under den gruppen, og de hadde fullføringstidspunkt i tidsrommet mellom 198012 og 198706. De som har fullførtdato etter 198509 er ugyldige og blir behandlet av datokontrollen.

BHU 1990 inneholder informasjon innhentet fra spesialundersøkelsen om utdanning tatt i utlandet per 1.okt. 1990, og inneholder derfor datoer langt tilbake i tid. BHU-1999 inneholder opplysninger om utdanninger tatt i utlandet per 1.10.1999. det vil derfor fra og med 1990 være en del gamle avslutningstidspunkt på BHU-filene.

De tidligste årgangene av BHU innhentet opplysninger fra Folke og Boligtellingene i 1970 og 1980 samt opplysninger fra filene over avsluttet utdannng. I 1991 og 1999 ble det i forbindelse med folke og bolig tellingene isteden gjennomført en spesialundersøkelse om utdanning fullført i utlandet. BHU har også siden 1992 blitt oppdatert med opplysninger fra Statens lånekasse for utdanning. Ved å benytte alternative kilder til avsluttede filene, kan man få en del fullføringstidspunkt som kan være langt tilbake i tid.

2.2.4. Endring og korreksjon på variabler

Ikke alle variablene på grunnlagsfilene egner seg til å bli lastet rett inn i NUDB. Flere av dem har vært utsatt for korreksjoner, - enten fordi de er oppdaget feil verdier på dem, eller at de må bearbeides for at vi skal få ut ønsket informasjon i databasen. Nedenfor gjøres en gjennomgang av hva som er gjort med ulike variabler på igangværende- og avsluttafilene. Noe av dette er beskrevet ovenfor i gjennomgangen av de ulike filenes beskrivelse, men her gjøres en mer detaljert gjennomgang av hva som er gjort.

2.2.4.1. Igangværende filene:

- Variablene 'Skolesla', 'Naring' eller 'Naering' er kodet om til henholdsvis 'SN72', 'SN78' eller 'SN94'. Hvordan navneforandringen blir, avhenger av filårgangen. Årgangene 1974-77 får 'SN72', årgangene 1978-1993 får 'SN78', mens årgangene f.o.m. 1994 får 'SN94'.
- Variabelen 'Skonr' har endret navn til 'Inr'.
- På bakgrunn av SN-variabelen er det laget den nye variabelen 'Univ'. Alle som har verdi '80' på enten 'SN72' eller 'SN78' har fått verdien '1' på 'Univ'. Alle som har verdien '80.301' på 'SN94' har fått verdien '1' på 'Univ'. Her er altså næringskoden (SNXX) benyttet til å identifisere dem som studerer på universitetet. Det er også lagd 'Univ' = '2'. Dette er records fra høyskoler som behandles som records fra universitetene ved oppbygning av NUDB, med unntak av setting av dato som skal være YYYY08. Disse er laget ved hjelp av 'Kilde'. Dersom 'Kilde' = '41', så er

'Univ' = '2'. Variabelen 'Univ' er senere benyttet til å avgjøre startdato. Startdato er satt til august for universitetsstudenter.

- Variabelen 'Utfall' er satt til blank på samtlige filer.
- Variabelen 'Kommnr' har endret noen verdier. Verdiene 'blank', '1888', '8000' og '9000' er satt til '9999', mens verdien '0300' er satt til '0301'.
- Variabelen 'Skolekom' har endret noen verdier. Verdiene 'blank', '1888', '8000' og '9000' er satt til '9999', mens verdien '0300' er satt til '0301'.
- Variabelen 'Heldel' har verdiene '1' og '2'. Alle andre verdier er satt til '1'.
- Variabelen 'Voksen' har verdiene '1' og '2'. Alle andre verdier er satt til '2'.
- For samtlige universitetsstudenter frem til 1995 på igangværendefilene er den opprinnelige NUS-koden de er satt opp med, erstattet av NUS-koden for fag i semesteret der denne er annerledes enn den opprinnelige på igangværendefilen. Dette er gjort ved å koble igangværendefilene mot de igangværendefilene som har registrert fag i semesteret. Her er fødselsnummer og NUS-kode benyttet som koblingsnøkkel. Ikke alle årganger har tilsvarende fag i semesteret filer. Disse har derfor beholdt sin opprinnelige NUS-kode.
- 'Startdato', dvs. når eleven/studenten påbegynte den gjeldende utdanningen, er laget ut ifra enten 'Estart' (elevens starttidspunkt) eller 'Kstart' (Kursets starttidspunkt). Disse variablene er på formatet: 'YYMM' (f.o.m. 1999-årgangen er den på YYYYMM). I tillegg er det derfor påført århundre, slik at 'Startdato' er på følgende format: 'YYYYMM'. 'Startdato' er primært laget av 'Estart', men dersom denne er blank, er 'Kstart' benyttet. Hvor datoene er hentet fra, er merket i 'Mrk_dl'. Elever/studenter som har startdato senere enn oktober for det året filen gjelder (f.eks. 199910) er slettet da dette er ugyldige startdatoer. Noen har også fått laget startdato om denne er blank på både 'Estart' og 'Kstart'.
- Variablene 'Studretn' og 'Kurstrin' er laget på bakgrunn av variabelen 'Studret'. 'Studretn' er laget på bakgrunn av første og andre siffer i 'Studret', mens 'Kurstrin' er laget på bakgrunn av dens femte siffer.
- Alle filer er påsatt variabelen 'Aargang' som forteller om når filen er gyldig for, dvs. det året grunnlagsfilen er fra. Format: 'YYYYMM'.
- Siden årgangene 1974-86 kun har NUS73, er disse pålagt NUS89 og klassetrinnene som gjelder for NUS89. Dette er gjort ved å benytte omkodingskatalog mellom NUS73 og NUS89.
- NUS2000 skal være med i miniregistrene. De igangværende årgangene som mangler NUS2000 er derfor påført denne, samt klassetrinn for NUS2000. I tillegg er variablene 'Kodetype', som forteller noe om dette er en enkeltutdanning eller en samlekode, og 'Uhgruppe', som forteller noe om hvilken universitets- eller høgskoleutdanning dette er, pålagt.
- Den internasjonale utdanningskoden 'ISCED97', samt dens fire tilleggsdimensjoner er også lagt på med NUS2000 som nøkkel. Hver NUS2000- kode har en korresponderende ISCED97-kode som også er lagt på filen.

På bakgrunn av første siffer i 'NUS2000' er variabelen 'Hoved' laget med følgende verdier:

- '1' dersom første siffer i NUS2000 er 0,
- '2' dersom første siffer i NUS2000 er 1 eller 2,
- '3' dersom første siffer i NUS2000 er 3 eller 4,
- '4' dersom første siffer i NUS2000 er 6, 7 eller 8, og
- '9' dersom første siffer i NUS2000 er 9

- Variabelen 'Ktrinn' (Gruppering av kurstrinn) er laget på bakgrunn av hvilke verdier recorden har på 'Kurstrin' og 'Klassetrinn2000'. Variabelen har fire verdier ('1', '2', '3' og '4').
- 760 Ingeniørstudenter som hadde nus89-koden '419101' har blitt endret til '419113' på igangværende i 1997. Samtidig har de fått ny 'studieretning' (21), ny 'ktype' (P) og ny 'komp' (1).
- Variabelen 'Utd' lages på bakgrunn av en rekke kriterier og variabler. Følgende regler er benyttet for å lage de ulike verdiene på årgangene:

For årgangene 1989 t.o.m. 1992:

Hvis første siffer i NUS2000 = 3, 4, 5 og utland = 1 så 'Utd' = 610 Videregående utdanning i utlandet

Hvis første siffer i NUS2000 > 5 og utland = 1 så 'Utd' = 620 Høyere utdanning i utlandet.

Hvis første siffer i NUS2000 = 2 **Og** 'Klbet' er ulik 000 **Og** 'Inr' ikke starter på 999 **Og** SN78 < 30, **Så** 'Utd' = 100 Grunnskoler

Hvis NUS2000 er lik 301104 **Eller** 301105 **Eller** 301106 **Så** 'Utd' = 510 Folkehøgskoler

Hvis 'Studret' = F **Så** 'Utd' = 212 Lærlinger/fagprøver

Hvis 'Klbet' = 000 **Og** 'Inr' starter på 999 **Så** 'Utd' = 520 Arbeidsmarkedskurs

Hvis 'Studretn' ikke er blank **Og** 'Klbet' ulik 000 **Og** 'Inr' ikke starter på 999 **Og** 'Voksen' = 2, **Så** 'Utd' = 211 Videregående skoler

Hvis første siffer i NUS2000 > 5 **Og** SN78 er ulik 81 **Så** 'Utd' = 399 Høgskoler

Hvis 'Univ' = 81 **Så** 'Utd' = 400 Universiteter og vitenskapelige høgskoler

Ellers 'Utd' = 220 Annen videregående utdanning I tillegg har videregåen

For årgangene 1993-1998 (Senere årganger har variablene liggende på årgangsfilene):

Hvis første siffer i NUS2000 = 3, 4, 5 og utland = 1 så 'Utd' = 610 Videregående utdanning i utlandet

Hvis første siffer i NUS2000 > 5 og utland = 1 så 'Utd' = 620 Høyere utdanning i utlandet.

Hvis første siffer i NUS2000 = 2 **Og** ('Klbet' er ulik 000 **Og** 'Inr' ikke starter på 999) **Og** SN94 < 80.200 **Så** 'Utd' = 100 Grunnskoler

Hvis NUS2000 er lik 301104 **Eller** 301105 **Eller** 301106 **Så** 'Utd' = 510 Folkehøgskoler

Hvis 'Studret' = F **Så** 'Utd' = 212 Lærlinger/fagprøver

Hvis 'Klbet' = 000 **Og** 'Inr' starter på 999 **Så** 'Utd' = 520 Arbeidsmarkedskurs

Ellers hvis 'Studretn' ikke er blank **Så** 'Utd' = 211 Videregående skoler

Hvis SN94= 80.302 **Så** 'Utd' = 311 Statlige høgskoler

Hvis SN94= 80.303 **Så** 'Utd' = 312 Militære høgskoler

Hvis SN94= 80.309 **Så** 'Utd' = 313 Andre høgskoler

Hvis 'Univ' = 81 **Så** 'Utd' = 400 Universiteter og vitenskapelige høgskoler

Ellers 'Utd' = 220 Annen videregående utdanning

- Variabelen 'Komp' er lagt på igangværende for 1997 og 1998, samt avslutta for 1997/1998. Denne er lagt på ved å matche NUS89 mot NUS89-katalogen (\$UTD/nuskat/arkiv/NUS89/g2000.dat). Senere årganger har variabelen liggende på årgangsfilen. Siden variabelen ikke er datert i NUDB kan den ikke brukes før 2000

2.2.4.2. Avsluttafilene:

På avsluttafilene er det gjort det samme som er nevnt i punktene under igangværende utdanning. Punktene nedenfor gjelder spesielt for avsluttafilene og kommer i tillegg:

- På avsluttafilene er det tatt med 'Sluttdato'. Denne er beregnet ut ifra 'Eslutt', eller 'Kslutt' dersom 'Eslutt' er blank. Hvilke kombinasjoner av 'Startdato' og 'Sluttdato' som er benyttet er merket i 'Mrk_dl1'. 'Startdato' som er satt etter filens (september) gyldighetsområde slettes. Det samme gjøres med 'Sluttdato' som er satt før (oktober) filas gyldighetsår.
- På avsluttafilene på 1980-tallet viser det seg å være relativt mange med sluttdato før filens gyldighetsområde. Dette er vesentlig personer med avbrutt grunnskole. De få som har utfall 8 (fullført) har kommet med i NUDB via BHU-filene dersom de har ført til oppdatering av BHU.
- Ugyldige datoer er koblet mot records på igangværende filene med fødselsnummer og NUSXX som nøkkel. På denne måten har vi greid å gi en del av de ugyldige recordene gyldige datoer. I tillegg er det også satt datoer kunstig hvor disse er feil, og heller ikke finnes på igangværendefilene. Kunstige datoer er satt til samme år som filen er gyldig for. For deleksamener som ikke har funnet makker ved kobling mot igangværendefilen er startdato satt til begynnelsen av det semesteret deleksamenen ble avsluttet i.
- Variabelen 'Ltype' er lagt på avsluttafilen 1997/1998 ved å matche avsluttafilen med \$UTD/wk01/sos9798.sas7bdat vha. nøkkelen 'fnr' og 'NUS89'. De som matcher har fått lagt på 'Ftype'...fra \$UTD/wk01/sos9798.sas7bdat
- På avsluttafilen 1975/76 er det ingen som har startdato 197609. Det har heller ikke årgangsfilen

2.2.4.3. BHU-filene:

Ikke alle variablene på grunnlagsfilene har egnet seg for å bli lastet rett inn i NUDB. En del variabler er blitt korrigeret av oss, - enten fordi det er oppdaget feil verdier på dem, eller fordi de måtte bearbeides for at vi skulle få ut ønsket informasjon fra NUDB. Nedenfor gjøres en gjennomgang av hva som er gjort med ulike variablene på BHU-filene. Noe av dette er beskrevet punkter ovenfor i gjennomgangen av BHU-filenes beskrivelse, men her gjøres en mer detaljert gjennomgang av hva som er gjort.

- Siden årgangen 1970 kun hadde NUS73 opprinnelig, er den påført NUS89 og de klasstrinnene som gjelder for NUS89. Dette er gjort ved å lage en omkodingskatalog mellom NUS73 uomkodet som finnes på filen, og NUS89 omkodet som finnes på BHU-1980 filen. Omkodningene på årgangene 1980-1986 var allerede foretatt da prosjektet med NUDB startet. De som ligger på BHU-1980 med fullføringstidspunkt 197011 kommer fra BHU-1970, og disse har fått utdanningskoden omkodet til NUS89-standarden. Vi benyttet derfor dette som utgangspunkt, og matchet BHU-1980 og BHU-1970 på fødselsnummer og fjernet dubletter med samme NUS72 og NUS89. Denne filen omfatter NUS72 og NUS89, samt klasstrinn, og danner starten på en omkodingskatalog mellom NUS72 og NUS89. Manglende koder ble lagt inn i katalogen manuelt. En del av NUS73u-kodene (u-omkodet) gikk til forskjellige NUS89o-koder (omkodet), men siden disse har forskjellig klasstrinn så dette greit. NUS73u-koder på samme klasstrinn som går til forskjellige NUS89o-koder ble behandlet ved at noen fødselsnummer ble sjekket på inn-filene. Antallet ble også sjekket. Det var ikke så mange som fikk feil kode på dette viset. Vi valgte da den mest riktige koblingen mellom NUS73u-NUS89o både etter fag og antall. De ukorrekte kodene ble fjernet, og vi sto igjen med en omkodingskatalog med 871 kombinasjoner. Omkodingskatalogen er lagret på Unix under: \$UTD/nuskat/arkiv/omkod_nus73u_nus89o/omk73u_89o.dat. Se for øvrig: Q:\DOK\Utdstat\Utd_standarder\Nus89\Omkod_Nus73_Nus89, og X:\360\Utdnivaa\Kataloger\NUS\omkod_nus73_nus89\Omkodkatalog_nus1973u-nus1989o.
- Variabelen 'Kommnr' har fått endret noen kodeverdier. Kodeverdiene 'blank', '0000', '1888', '2111', '3030', '8000' og '9000' er satt til '9999' (uoppgift). Koden 1249 som var på BHU-1985 var ikke

gyldig kode for dette året og ble satt til 1201. I tillegg har en del sammenslåtte kommuner blitt kodet korrekt i de respektive årgangene etter følgende liste:

Kommunesammenslåinger fra 1/1-1988 til 1/1-2008

Endringstidspunkt	Fra kommune	Ny kommuneenhet
01.01.2008	1503 Kristiansund	1505 Kristiansund
01.01.2008	1556 Frei	1505 Kristiansund
01.01.2006	1154 Vindafjord	1160 Vindafjord
01.01.2006	1159 Ølen	1160 Vindafjord
01.01.2006	1569 Aure	1576 Aure
01.01.2006	1572 Tustna	1576 Aure
01.01.2005	1804 Bodø	1804 Bodø
01.01.2005	1842 Skjerstad	1804 Bodø
01.06.2002	0701 Borre	0701 Horten
01.01.2002	0716 Våle	0716 Re
01.01.2002	0718 Ramnes	0716 Re
01.01.2002	1214 Ølen	1159 Ølen
01.01.1994	0103 Fredrikstad	0106 Fredrikstad
01.01.1994	0113 Borge	0106 Fredrikstad
01.01.1994	0131 Rolvsøy	0106 Fredrikstad
01.01.1994	0133 Kråkerøy	0106 Fredrikstad
01.01.1994	0134 Onsøy	0106 Fredrikstad
01.01.1992	2001 Hammerfest	2004 Hammerfest
01.01.1992	2016 Sørøysund	2004 Hammerfest
01.01.1992	0903 Arendal	0906 Arendal
01.01.1992	0918 Moland	0906 Arendal
01.01.1992	0920 Øyestad	0906 Arendal
01.01.1992	0921 Tromøy	0906 Arendal
01.01.1992	0922 Hisøy	0906 Arendal
01.01.1992	0401 Hamar	0403 Hamar
01.01.1992	0414 Vang	0403 Hamar
01.01.1992	0102 Sarpsborg	0105 Sarpsborg
01.01.1992	0114 Varteig	0105 Sarpsborg
01.01.1992	0115 Skjeberg	0105 Sarpsborg
01.01.1992	0130 Tune	0105 Sarpsborg
01.01.1988	0707 Larvik	0709 Larvik
01.01.1988	0708 Stavern	0709 Larvik
01.01.1988	0725 Tjølling	0709 Larvik
01.01.1988	0726 Brunlanes	0709 Larvik
01.01.1988	0727 Hedrum	0709 Larvik
01.01.1988	0705 Tønsberg	0704 Tønsberg
01.01.1988	0721 Sem	0704 Tønsberg
01.01.1988	0703 Horten	0701 Borre
01.01.1988	0717 Borre	0701 Borre

- På BHU-1993, 1994, 1995, 1996 og 1997 er NUS2000- koden '616101' korrigeret til NUS2000 '516101'. Denne har NUS89-kode 628901, og var plassert på nivå 6 i NUS2000, men er ikke godkjent av KUF. Er derfor flyttet til nivå 5. NUS2000-kode 616101 er fjernet.
- Ut fra fødselsnummeret er det laget to nye hjelpevariabler, henholdsvis 'Fodt_aar' (YYYY) og 'Fodt_dato' (DDMMYYYY). Disse variablene er ikke lagt inn i basen, men er benyttet for å gjennomføre en kontroll av om det finnes elever/studenter som er under 16 år i BHU-registrene. Testen tar også hensyn til at noen kan være født på 1800-tallet. Dette gjøres ved å finne personer som har verdier mellom 500 og 749 som 7, 8 og 9 siffer i fødselsnummeret.
- Alle filer har blitt påført variabelen 'aargang'. Denne forteller hvilket år filen er gyldig for, mao. det året grunnlagsfilen er fra.

- NUS2000 skal være med i miniregistrene. BHU-årgangene fra og med 1970 til og med 1998 som manglet NUS2000 har derfor fått koblet på dette kjennemerket, samt klasstrinn for NUS2000. I tillegg er variablene 'kodetype', som forteller noe om dette er en enkeltutdanning eller en samlekode, og 'uhgruppe', som forteller noe om hvilken universitets- eller høgskoleutdanning dette er, påført. Dette ble gjort ved å benytte en omkodingskatalog mellom NUS89 og NUS2000. Omkodingskatalogen ligger lagret under www.ssb.no/emner/04/90/nos_c617/

- På bakgrunn av første siffer i NUS2000 (nivå) lages variabelen 'hoved'. Variabelen har fem verdier:

- 1 = nivå '0' i NUS2000 (ingen utdanning)
- 2 = nivå '1' og '2' i NUS2000 (obligatorisk utdanning)
- 3 = nivå '3', '4' og '5' i NUS2000 (mellomutdanning)
- 4 = nivå '6', '7' og '8' i NUS2000 (universitets- og høgskoleutdanning)
- 9 = nivå '9' i NUS2000 (uoppgitt utdanning)

- Variabelen 'kildebhu' opptrer først på BHU-1999, og er derfor kodet på de tidligere årgangene. Kodelisten for denne variabelen er som følger:

- 01 = 'Fra Folke- og boligtellingsen 1970'
- 02 = 'Fra avsluttet-filen'
- 03 = 'Fra Statens lånekasse for utdanning'
- 04 = 'Fra spesialundersøkelsen om skolegang i 1999'
- 05 = 'Fra Folke- og boligtellingsen 1980'
- 06 = 'Tilganger fra FS-skoler på BHU-2000'
- 99 = 'Uoppgitt'

Vi vet at alle på BHU-1970, samt alle på BHU-1980 med fullførtdato 197011 er fra Folke- og boligtellingsen i 1970, og at de på BHU-1980 med opprinnelig fullføringstidspunkt 198011 er fra Folke- og boligtellingsen i 1980. Årgangene 1985-1998 ble satt til koden for uoppgitt. Denne variabelen ble innført på BHU-1999, da det nå eksisterer flere ulike kilder til BHU-opplysninger. Det kan også bli aktuelt å innhente informasjon fra en rekke andre registre i tiden framover, slik at denne variabelen kan få økt nytte.

- Variabelen 'pedsem', var ikke med på BHU-1970 og ble derfor kodet på grunnlag av NUS2000. De som hadde NUS2000 kodene '623801' og '623802', fikk koden '1' som står for 'Fullført praktisk-pedagogisk utdanning, resten fikk stå 'blank' som står for 'Uten praktisk-pedagogisk utdanning'. Variabelen var allerede kodet på resten av årgangene.
- I NUDB ønsket vi at ingen skulle stå med 'blank' på variabelen 'kltrinn2000' (klasstrinn etter NUS2000-standard). Fra og med BHU 1990 manglet imidlertid en rekke NUS2000 koder klasstrinn. Vi vurderte å påføre det høyeste klasstrinnet til NUS2000 koden fra NUS2000 katalogen. Dette ville imidlertid ikke bli helt korrekt, og vi valgte derfor å la de blanke klasstrinnene stå blanke isteden for å korrigere dem til noe vi ikke med sikkerhet kunne si ble korrekt. Det ble lagt inn programmene at NUS2000 kodene '099901' og '999999' skulle få påført klasstrinn '00'. Vi la inn en liten sjekk på om det fantes noen ugyldige klasstrinn på BHU-filene. Dette ble undersøkt ved å lage en krysstabell med 'NUS2000' og 'kltrinn2000'. Records med klasstrinn '55' og NUS2000 '329999', ble korrigeret til klasstrinn '12'.

- På miniregister BHU 1990 er 28.625 klasstrinn blanke
- På miniregister BHU 1991 er 27.258 klasstrinn blanke
- På miniregister BHU 1992 er 25.613 klasstrinn blanke
- På miniregister BHU 1993 er 33.116 klasstrinn blanke

- På miniregister BHU 1994 er 32.948 klasstrinn blanke
 - På miniregister BHU 1995 er 34.141 klasstrinn blanke
 - På miniregister BHU 1996 er 34.539 klasstrinn blanke
 - På miniregister BHU 1997 er 35.343 klasstrinn blanke
 - På miniregister BHU 1998 er 36.629 klasstrinn blanke
 - På miniregister BHU 1999 er 35.825 klasstrinn blanke
- BHU-filene fra og med 1970 til og med 1989 manglet statsborgerskap. Vi har kun klart å finne opplysninger om dette fra og med 1986. Statsborgerskap ble koblet på disse filene via ferdige årgangsfiler for befolkningen. Filene finnes fra og med 1986 og er lagret som flate filer på: \$BESTAT/folkem/arkiv/bosatt/g1986m01d01 - g1990m01d01.dat. BHU-filene er per 1.10 og befolkningsfilene fra 1.1. Vi valgte å benytte fila fra 1.1.1987 for å påføre statsborgerskap på BHU-1986, 1.1.1988 for å påføre BHU-1987 osv. Dette gjorde vi for at det da kun ble 3 måneder forskjell på datoen når vi innhenter opplysninger om statsborgerskap, imot 7 måneder om vi hadde benyttet filen per 1.1.1986 til BHU-1986. Vi fikk imidlertid ikke treff på alle recordene og satte disse til koden '990' som står for uoppgitt. De årgangene som hadde noen records som var blanke ble satt til '990'. Mangler det opplysning for hele årgangen, settes hele årgangen til 'blank'.
 - I 1986 var det 15.074 som ikke fikk match på fødselsnummer og statsborgerskap i koblingen (3 var blanke og ble satt til '990').
 - I 1987 var det 58.192 som ikke fikk match på fødselsnummer og statsborgerskap i koblingen (19 var blanke og ble satt til '990').
 - I 1988 var det 60.325 som ikke fikk match på fødselsnummer og statsborgerskap i koblingen (18 var blanke og ble satt til '990').
 - I 1989 var det 22.944 som ikke fikk match på fødselsnummer og statsborgerskap i koblingen (ingen var blanke).
 - De som hadde blank på variabelen 'fulland' (fullføringsland for utdanningen) fikk isteden koden '990' for uoppgitt. Mangler det opplysning for hele årgangen, settes hele årgangen til 'blank'.
 - Variabelen 'uhgruppe' manglet ledende nuller i BHU-1998, og fikk påført dem.
 - Variabelen 'fulland' er blank til og med årgang 1989.
 - Ved sammenligning mot offisiell statistikk viste det seg at det var store forskjeller mellom BHU-1989 miniregister og tall fra Statistisk Årbok. Det viste seg at det var lagret to versjoner av BHU-1989 på stormaskinformat. Versjon 02 som lå overført til Unix hadde fått påført resultat fra spesialundersøkelsen "Utdanning fullført i utlandet", som ble gjennomført våren 1991, mens versjon 00 ikke hadde påført disse opplysningene. Undersøkelsen hentet inn opplysninger om eksamen avlagt i utlandet av utenlandsfødte personer som hadde innvandret for første gang mellom 1.11.1980 og 31.12. 1990. Det synes mest korrekt å benytte BHU-1989 uten disse opplysningene, da de er påført BHU-1990

3. Tiltrettelegging av miniregistrene

3.1. Variable på miniregistre

Med utgangspunkt i grunnlagsregistrene er opprettet s.k. miniregistre. Dette er vår benevnelse på de korrigerede og omkodete registrene som er benyttet som input til NUDB. Miniregistrene inneholder et utvalg av variabler fra grunnlagsregistrene i tillegg til et sett av nye variabler som er kodet på grunnlag av gamle.

3.1.1. Igangværende- og avslutta utdanning

For å lette innlastingen i NUDB når forløpet dannes har vi valgt å gi miniregistrene for igangværende og avslutta utdanning samme filbeskrivelse. Det vil si at miniregistrene for disse inneholder de samme variablene, men at verdiene er blanke der det ikke eksisterer opplysninger om variabelen, eller der det ikke skal innhentes opplysninger om variabelen inn i databasen. Miniregistrene for igangværende- og avslutta utdanning har følgende variabler (variablene er angitt i den rekkefølgen de eksisterer på filene):

Variabelnavn	Variabelbeskrivelse	Type	Lengde	Format	Gjelder årgang
aargang	Årgang. Variabelen henviser til årgangen filen er gyldig for.	Num	4	YYYY	
fnr	Fødselsnummer	Char	11	Ledende nuller	
nus2000	Utdanningens art etter NUS2000-koden. For årganger yngre enn 1999, hentes NUS2000 inn fra omkodingskatalog mellom NUS89 og NUS2000.	Char	6		
nus89	Utdanningens art etter NUS89-koden.	Char	6		
kltrinn2000	Klassetrinn etter NUS2000.	Num	2		
heldel	Heltid/deltid. Sier noe om eleven/studenten går på skolen på hel- eller deltid. Variabelen har bare to verdier i miniregistrene, selv om den hadde flere verdier i grunnlagsregistrene. Heltid = 1, Deltid = 2.	Char	1		
startd	Startdato. Sier noe om når eleven/studenten påbegynte sin utdanning.	Char	6	YYYYMM	
sluttd	Sluttdato. Sier noe om når eleven/studenten avsluttet utdanningen sin.	Char	6	YYYYMM	
voksen	Voksenopplæring. Gir informasjon om studiet er en del av et voksenopplæringstilbud som går inn under lov om voksenopplæring. Voksenopplæring = 1, Ikke voksenopplæring = 2.	Char	1		
studretn	SSBs studieretningskode. Lages på bakgrunn av de to første sifrene i variabelen 'Studret'.	Char	2		
kurstrin	Kurstrinn. Lages på bakgrunn av femte siffer i variabelen 'Studret'.	Char	1		
komp	Elevenes kompetanse etter VKII/ VKIII. Fra 1997(kan ikke brukes før fra 2000)	Char	1		
realkomp	Realkompetanse. Denne variabelen får et merke dersom studenten er tatt opp på studiet på bakgrunn av realkompetanse, og hvilken kompetanse det dreier seg om. Fra 2001	Char	1		

Variabelnavn	Variabelbeskrivelse	Type	Lengde	Format	Gjelder årgang
uhgruppe	Gir informasjon om hvilken type universitets- eller høyskoleutdanning studenten er igangværende med eller har avslutta.	Char	2		
ktrinn	Gruppering av kurstrinn. Kode 1-4.	Char	1		
kilde	Kilde	Char	2		
rett	Rett	Char	1		
ftype	Lærlingetype	Char	1		
kodetype	Variabelen gir opplysninger om dette er en utdanning som faller inn som 'samlekode' eller 'enkeltutdanning' i NUS2000. Samlekode = 1, Enkeltutdanning = 2.	Char	1		
evufjern	Etter-, videreutdanning eller fjernundervisning. Variabelen gir opplysninger om eleven/studenten tar utdanningen som etter-, videreutdanning, eller som fjernundervisning.	Char	1		
i97isced	Dette er den internasjonale utdanningskoden som gir opplysninger om hvilken utdanning eleven/studenten tar.	Char	4		
i97destn	Dette er en tilleggskode til ISCED97.	Char	1		
i97orien	Dette er en tilleggskode til ISCED97.	Char	1		
i97grads	Dette er en tilleggskode til ISCED97.	Char	1		
i97varig	Dette er en tilleggskode til ISCED97.	Char	1		
utfall	Utfall omkodet. Variabelen inneholder opplysninger om utfallet på studiet. Avbrutt = 2, Fullført = 8. På årganger fra og med 1995, er det flere koder.	Char	1		
utfallu	Utfall uomkodet	Char	1		
rectype	Recordtype på avslutta utdanning.	Char	1		
studpoenggrad	Studiepoeng for grad.	Num	5.1		1999-
studpoengeks	Studiepoeng for eksamen.	Num	5.1		1999-
gjentak	Inneholder opplysninger om dette er et studium som gjenopptas.	Char	1		
studrett	Med eller uten studierett.	Char	1		
utd	Gruppering av skoleslag.	Char	3		
skolekom	Skolekommune.	Char	4		
sn72	Næringskode fra 1972. Filene fra 1974-1977 har denne.	Char	2		
sn78	Næringskode fra 1978. Filene fra 1978-1993 har denne.	Char	2		
sn94	Næringskode fra 1994. Filene fra 1994- har denne.	Char	6		
eierf	Skolens eierforhold.	Char	1		
hskode	Kode for høyskoler. Gir informasjon om hvilken høyskole studenten studerer ved.	Char	3		
inr	Identifikasjonsnummer. (skolenummer fram til 1998)	Char	7		
orgnr	Organisasjonsnummer. (skolenummer fra 2000)	Char	9		
sslagn	Skoleslag, omkodet.	Char	1		
fulland	Fullføringsland.	Char	3		
univ	Merkevariabel som indikerer spesiell behandling ved innlegging i basen	Char	1		

Variabelnavn	Variabelbeskrivelse	Type	Lengde	Format	Gjelder årgang
hoved	Hovedgruppe-variabel. Avledet av NUS2000	Char	1		
mrk_fnr	Merking av fødselsnummer.	Char	1		
mrk_dl	Merking av startdato.	Char	1		
mrk_dl1	Merking av start- og sluttdato.	Char	1		
mrk_dl2	Merking av start- og sluttdato.	Char	1		
mrk_dl3	Merking av start- og sluttdato.	Char	1		
mrk_per	Merking av fødselsnummer.	Char	1		
mrk_end	Merking av startdato.	Char	1		
snr_nudb	S-nummer	Char	7		
mrk_snr_nudb	Merking av S-nummer	Char	1		
mrk_grskole	Merking av grunnskole	Char	1		
spesund	Spesialundervisning etter enkeltvedtak	Char	1		2002-
morsmal	Støtteundervisning i morsmål	Char	1		2002-
sprakopp	Særskilt norskopplæring for språklige minoriteter	Char	1		2002-
doktype	Dokumentasjonstype	Char	2		2002-
utland	Merke for utenlandsstudenter	Char	1		

3.1.2. BHU

Miniregisterene for BHU er filer med de variablene vi har tatt ut fra grunnlagsfila, samt variabler som er dannet i løpet av prosessen. For å lette innlastingen i databasen har vi valgt å gi samtlige miniregistrene for befolkningens høyeste utdanning lik filbeskrivelse. Det vil si at miniregistrene for alle årgangene inneholder de samme variablene, men at verdiene er blanke hvis det ikke kan innhentes opplysninger om variabelen inn i databasen.

Miniregistrene for BHU som skal brukes til ilegg i basen ligger lagret på Unix som tekstfiler under følgende stier:

```
$UTD/nudb/arkiv/bhu/g1970.txt
$UTD/nudb/arkiv/bhu/g1980.txt
$UTD/nudb/arkiv/bhu/g1985.txt
$UTD/nudb/arkiv/bhu/g1986.txt
$UTD/nudb/arkiv/bhu/g1987.txt
$UTD/nudb/arkiv/bhu/g1988.txt
$UTD/nudb/arkiv/bhu/g1989.txt
$UTD/nudb/arkiv/bhu/g1990.txt
$UTD/nudb/arkiv/bhu/g1991.txt
$UTD/nudb/arkiv/bhu/g1992.txt
$UTD/nudb/arkiv/bhu/g1993.txt
$UTD/nudb/arkiv/bhu/g1994.txt
$UTD/nudb/arkiv/bhu/g1995.txt
$UTD/nudb/arkiv/bhu/g1996.txt
$UTD/nudb/arkiv/bhu/g1997.txt
$UTD/nudb/arkiv/bhu/g1998.txt
$UTD/nudb/arkiv/bhu/g1998.spesu (BHU-1998 med påført spesialundersøkelsen om skolegang 1991)
$UTD/nudb/arkiv/bhu/g1999.txt
$UTD/nudb/arkiv/bhu/g2000.txt
```

Miniregistrene har følgende variabler:

Variabelnavn	Variabelbeskrivelse	Type	Lengde	Format
fnr	Fødselsnummer	char	11	
snr	S-nummer	char	8	
kommnr	Bostedskommune	char	4	
statborg	Statsborgerskap	char	3	
NUS89	Utdanningens art etter NUS89-standarden (omkodet)	char	6	
NUS2000	Utdanningens art etter NUS2000-standarden (omk.)	char	6	
kltrinn2000	Klassetrinn etter NUS2000-standarden (omk.)	char	2	
pedsem	Fullført praktisk-pedagogisk seminar	char	1	
uhgruppe	Gruppering av høyere utdanning	char	2	
i97iscd	ISCED97, (omk.)	char	1	
i97destn	ISCED97 Destination	char	1	
i97orien	ISCED97 Orientation	char	1	
i97grads	ISCED97 Gradsstruktur	char	1	
i97varig	ISCED97 Varighet	char	1	
kildebhu	Kilde for BHU-opplysning	char	1	
fulland	Fullføringsland	char	3	
fulldato	Fullføringstidspunkt (YYYYMM)	char	6	
hoved	Hovedgruppe-variabel	char	1	
kodetype	Kodetype	char	1	
mrk_fnr	Fødselsnummer-merking	char	1	
mrk_snr_nudb	S-nummer- merking	char	1	
mrk_dl_bhu	Merker omkodinger på fullføringstidspunktet	char	1	
aargang	Årgangsupplysning for miniregisteret	char	4	

3.2. Korreksjoner og kontroller

Her foretas en gjennomgang av statistikk på grunnlagsregistrene, hvor mange records som fjernes, og hvilke merkinger filene har fått på miniregistrene.

3.2.1. Kontroller på filene for igangværende utdanning per 1.okt. hvert år

Igangværende utdanning	1974	1975	1976	1977
Antall records på grunnlagsregistrene	221976	227979	229918	234884
Antall dubletter (fjernet)	660	781	950	936
Antall records som har startdato etter filas utløp (slettes)	-	-	29	1
Antall records i miniregisteret	221316	227198	228939	233947
<u>Utdanningsnivå (Miniregisteret):</u>				
Grunnskolenivå	9164	8093	5655	3963
Videregående skolenivå	144147	149243	153691	160814
Universitets- og høghskolenivå	68005	69862	69593	69170
Universitet	38971	40208	40009	38858
Herav som har skiftet verdi på utdanningens art(fra studiemål til fag i semesteret)	27606	27456	26860	25584
<u>Dato:</u>				
Antall records som mangler startdato	-	-	-	-
Antall records hvor elevens startdato er brukt (Mrk_dl = 1)	182345	205409	209331	216231
Antall records hvor kursets startdato er brukt (Mrk_dl = 3)	-	-	4503	7719
Antall records hvor uoppgitt startdato er satt (Mrk_dl = 5)	-	-	-	-
Antall records hvor startdato er satt til filas årgang (Mrk_dl = 6)	38971	21789	15105	9997
<u>Fødselsnummerkontroll:</u>				
Antall med gyldig fødselsnummer (Mrk_fnr = 0)	220798	226707	228368	233320
Antall med gyldig d-nummer (Mrk_fnr=1)	-	-	-	-
Antall med nuller i personnummer (Mrk_fnr = 2)	-	-	-	-
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	518	491	571	627

Igangværende utdanning	1978	1979	1980	1981
Antall records på grunnlagsregistrene	245653	253666	259131	268504
Antall dubletter (fjernet)	1078	1030	4	1235
Antall record som har startdato etter filas utløp (slettes)	-	-	26	-
Antall records i miniregisteret	244575	252636	259102	267269
<u>Utdanningsnivå (Miniregisteret):</u>				
Grunnskolenivå	3231	2856	2776	2685
Videregående skolenivå	169851	175218	179560	183900
Universitets- og høghskolenivå	74562	74562	76765	80684
Universitet	38912	40074	40039	39420
Herav som har skiftet verdi på utdanningens art	25102	26450	26115	25990
<u>Dato:</u>				
Antall records som mangler startdato	-	-	-	-
Antall records hvor elevens startdato er brukt (Mrk_dl = 1)	238243	249140	255219	266394

Antall records hvor kursets startdato er brukt (Mrk_dl = 3)	155	-	1928	-
Antall records hvor uoppgitt startdato er satt (Mrk_dl = 5)	-	-	-	-
Antall records hvor startdato er satt til filas årgang (Mrk_dl = 6)	6177	3496	1955	875

Fødselsnummerkontroll:

Antall med gyldig fødselsnummer (Mrk_fnr = 0)	243790	251974	258512	266739
Antall med gyldig d-nummer (Mrk_fnr = 1)	-	-	-	-
Antall med nuller i personnummer (Mrk_fnr = 2)	-	-	-	-
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	785	662	590	530

Igangværende utdanning	1982	1983	1984	1985
Antall records på grunnlagsregistrene	278873	290305	301634	300785
Antall dubletter (fjernet)	1346	1636	1910	1858
Antall record som har startdato etter filas utløp (slettes)	-	66	45	137
Antall records i miniregisteret	277527	288607	299684	298791

Utdanningsnivå (Miniregisteret):

Grunnskolenivå	3064	3778	3987	3768
Videregående skolenivå	189092	196971	204647	210023
Universitets- og høghskolenivå	85371	87858	91050	84999
Universitet	40351	40687	41715	40995
Herav som har skiftet verdi på utdanningens art	27051	27226	28362	27367

Dato:

Antall records som mangler startdato	-	-	-	-
Antall records hvor elevens startdato er brukt (Mrk_dl = 1)	276614	288363	299465	298704
Antall records hvor kursets startdato er brukt (Mrk_dl = 3)	-	-	-	-
Antall records hvor uoppgitt startdato er satt (Mrk_dl = 5)	-	-	-	-
Antall records hvor startdato er satt til filas årgang (Mrk_dl = 6)	913	244	219	87

Fødselsnummerkontroll:

Antall med gyldig fødselsnummer (Mrk_fnr = 0)	277054	288039	299022	297536
Antall med gyldig d-nummer (Mrk_fnr = 1)	-	-	-	-
Antall med nuller i personnummer (Mrk_fnr = 2)	-	1	-	-
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	473	567	662	1255

Igangværende utdanning	1986	1987	1988	1989
Antall records på grunnlagsregistrene	312021	304943	319135	365306
Antall dubletter (fjernet)	3185	12	2	1
Antall record som har startdato etter filas utløp (slettes)	10	-	13	51
Antall records med utenlandsutdanning	6114	6472	6949	7183
Antall records i miniregisteret	314940	311403	325869	372447
<u>Utdanningsnivå (Miniregisteret):</u>				
Grunnskolenivå	2852	2669	2082	2469
Videregående skolenivå	207429	200954	208003	239522
Universitets- og høgskolenivå	104659	107780	115784	130456
Universitet	42469	43627	47311	56169
Herav som har skiftet verdi på utdanningens art	28093	43627	32115	40310
<u>Dato:</u>				
Antall records som mangler startdato	-	-	-	-
Antall records hvor elevens startdato er brukt (Mrk_dl = 1)	308700	70408	77568	88616
Antall records hvor kursets startdato er brukt (Mrk_dl = 3)	-	233551	240576	275342
Antall records hvor uoppgitt startdato er satt (Mrk_dl = 5)	6114	6472	6749	7183
Antall records hvor startdato er satt til filas årgang (Mrk_dl = 6)	402	1168	1201	1593
<u>Fødselsnummerkontroll:</u>				
Antall med gyldig fødselsnummer (Mrk_fnr = 0)	314182	310708	325179	371525
Antall med gyldig d-nummer (Mrk_fnr = 1)	-	-	-	-
Antall med nuller i personnummer (Mrk_fnr = 2)	1	1	-	1
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	1033	890	915	1208
Igangværende utdanning	1990	1991	1992	1993
Antall records på grunnlagsregistrene	385474	401231	414510	421507
Antall dubletter (fjernet)	2	2	7	4
Antall record som har startdato etter filas utløp (slettes)	164	126	191	125
Antall records med utenlandsutdanning	7491	7722	7906	7905
Antall records i miniregisteret	392799	408825	422218	429283
<u>Utdanningsnivå (Miniregisteret):</u>				
Grunnskolenivå	2376	2631	2812	3005
Videregående skolenivå	249888	256088	257267	253268
Universitets- og høgskolenivå	140535	150106	162139	173010
Universitet	62734	67372	72909	77027
Herav som har skiftet verdi på utdanningens art	45462	49327	52571	55325
<u>Dato:</u>				
Antall records som mangler startdato	-	-	-	-
Antall records hvor elevens startdato er brukt (Mrk_dl = 1)	96459	110733	120983	129621
Antall records hvor kursets startdato er brukt (Mrk_dl = 3)	287511	290226	293288	291710
Antall records hvor uoppgitt startdato er satt (Mrk_dl = 5)	7491	7722	7906	7905
Antall records hvor startdato er satt til filas årgang (Mrk_dl = 6)	1612	467	495	531
<u>Fødselsnummerkontroll:</u>				
Antall med gyldig fødselsnummer (Mrk_fnr = 0)	392052	408052	421610	428683
Antall med gyldig d-nummer (Mrk_fnr = 1)	1	-	-	-
Antall med nuller i personnummer (Mrk_fnr = 2)	2	1	-	9
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	1018	1095	1062	1075

Igangværende utdanning	1994	1995	1996	1997
Antall records på grunnlagsregistrene	414425	412415	417405	414559
Antall dubletter (fjernet)	2	5	1	2
Antall record som har startdato etter filas utløp (slettes)	109	228	32	5
Antall records med utenlandsutdanning	8363	9318	10471	11649
Antall records i miniregisteret	422677	421490	427843	426201
<u>Utdanningsnivå (Miniregisteret):</u>				
Grunnskolenivå	2874	2680	1939	1058
Videregående skolenivå	242467	232771	233675	232530
Universitets- og høgskolenivå	177336	186039	192229	192613
Universitet	79509	82957	85173	83484
Herav som har skiftet verdi på utdanningens art	57256	59926	-	-
<u>Dato:</u>				
Antall records som mangler startdato	-	-	-	-
Antall records hvor elevens startdato er brukt (Mrk_dl = 1)	138416	154431	163387	167035
Antall records hvor kursets startdato er brukt (Mrk_dl = 3)	275879	256211	253970	247510
Antall records hvor uoppgitt startdato er satt (Mrk_dl = 5)	8363	9318	10471	11649
Antall records hvor startdato er satt til filas årgang (Mrk_dl = 6)	467	2043	593	592
<u>Fødselsnummerkontroll:</u>				
Antall med gyldig fødselsnummer (Mrk_fnr = 0)	421921	420506	426902	424940
Antall med gyldig d-nummer (Mrk_fnr = 1)	1	-	3	2
Antall med nuller i personnummer (Mrk_fnr = 2)	13	7	4	2
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	1190	1490	1512	1842

Igangværende utdanning	1998	1999	2000	2001
Antall records på grunnlagsregistrene	414178	414304	411628	412763
Antall dubletter (fjernet)	2	-	-	-
Antall record som har startdato etter filas utløp (slettes)	2	145	136	0
Antall records med utenlandsutdanning	12680	13916	14832	15762
Antall records i miniregisteret	426854	427778	426319	428212
<u>Utdanningsnivå (Miniregisteret):</u>				
Grunnskolenivå	585	460	485	556
Videregående skolenivå	229907	223472	221635	216376
Universitets- og høgskolenivå	196362	203846	204199	211280
Universitet	81128	78969	81560	-
Herav som har skiftet verdi på utdanningens art	-	-	-	-
<u>Dato:</u>				
Antall records som mangler startdato	-	-	-	-
Antall records hvor elevens startdato er brukt (Mrk_dl = 1)	181929	298444	289413	287872
Antall records hvor kursets startdato er brukt (Mrk_dl = 3)	232238	-	-	-
Antall records hvor uoppgitt startdato er satt (Mrk_dl = 5)	12680	129334	136906	139725
Antall records hvor startdato er satt til filas årgang (Mrk_dl = 6)	591	827	679	923
<u>Fødselsnummerkontroll:</u>				
Antall med gyldig fødselsnummer (Mrk_fnr = 0)	425864	425654	423812	425178
Antall med gyldig d-nummer (Mrk_fnr = 1)	6	731	-	-
Antall med nuller i personnummer (Mrk_fnr = 2)	4	13	-	-
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	1564	2207	3186	3342

Igangværende utdanning	2002	2003	2004	2005
Antall records på grunnlagsregistrene				
Antall dubletter (fjernet)				
Antall record som har startdato etter filas utløp (slettes)				
Antall records med utenlandsutdanning	15347	15209	14371	13663
Antall records i miniregisteret	448277	464495	457037	467236
<u>Utdanningsnivå (Miniregisteret):</u>				
Ingen utdanning og førskoleutdanning (0-2)				-
Obligatorisk utdanning (1-2)	499	1498	650	699
Videregående, grunnutdanning (3)	138415	147903	147235	150543
Videregående, avsluttende utdanning (4)	76174	80751	75887	79829
Påbygging til videregående utdanning (5)	6160	6071	5455	8837
Universitet- og høyskoleutdanning, lavere nivå (6)	180366	177957	172665	170500
Universitet- og høyskoleutdanning, høyere nivå (7)	42234	45666	50072	51507
Forskerutdanning (8)	4429	4649	4632	5321
Uoppgitt nivå (9)	-	-	-	-
<u>Dato:</u>				
Antall records som mangler startdato				
Antall records hvor elevens startdato er brukt (Mrk_dl = 1)	302239	238847	307490	314183
Antall records hvor kursets startdato er brukt (Mrk_dl = 3)	-	-	-	-
Antall records hvor uoppgitt startdato er satt (Mrk_dl = 5)	146038	225648	149106	153053
Antall records hvor startdato er satt til filas årgang (Mrk_dl = 6)	-	-	-	-
<u>Fødselsnummerkontroll:</u>				
Antall med gyldig fødselsnummer (Mrk_fnr = 0)	443679	459724	453778	463814
Antall med gyldig d-nummer (Mrk_fnr = 1)	-	-	19	-
Antall med nuller i personnummer (Mrk_fnr = 2)	-	-	2	-
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	4598	4771	2797	3422
Gyldig fnr. med snr	443673	459720	452638	463023
Ugyldig fnr. Uten snr	4604	4775	3958	4213

3.2.2. Kontroller på filene for avslutta utdanning i skoleåret.

mrk_dl1 Merking av start- og stoppdato

Kodeliste:

A = Merkes hvis startd >= sluttd og Mrk_dl1 er 2 eller 4

B = andre kombinasjoner enn de som står over

1 = Henter både start og sluttdato fra estart og eslutt

2 = henter estart og kslutt

3 = henter kstart og eslutt

4 = henter kstart og kslutt

6 = henter kun kslutt

7 = merkes hvis startd =000000 eller 190001 og Mrk_dl1 = 1 eller 3

8 = merkes hvis startd =000000 eller 190001 og Mrk_dl1 = 2 eller 4

9 = Merkes hvis startd >= sluttd og Mrk_dl1 er 1 eller 3

mrk_dl2 Merking av start- og stoppdato

Kodeliste:

0 = Startd er satt

1 = Hvis match mot igangværende fila med fødselsnummer + utdanningens art som nøkkel så startd hentes fra igangværende fila

2 = Hvis match mot vårsemesterfila med fnr + utdanningens art som nøkkel så startd settes til start i vårsemesteret (YY01)

3 = Hvis match mot igangværende med fnr som nøkkel så startd settes til YY-1 09

4 = Startd settes til start i det semesteret kurset er avsluttet med

mrk_dl3 Merking av start- og stoppdato

Kodeliste:

0 = Hvis start- og sluttdato (dvs. alle variabler) er hentet fra samme record

1 = Hvis startdato, men ikke sluttdato, hentes fra den recorden hvor de andre opplysningene er hentet fra

2 = Hvis stoppdato, men ikke startdato, hentes fra den recorden hvor de andre opplysningene er hentet fra

3 = Hvis start- og stoppdato hentes fra en record mens de andre opplysningene hentes fra en annen record

4 = Hvis startdato, stoppdato og de andre opplysningene hentes fra tre ulike records

Avslutta	1975	1976	1977	1978
Antall records på grunnlagsregistrene	234568	231368	230914	234088
Antall records på grunnlagsregisteret med utfall=8	207467	202127	201727	204185
<u>Datomerking¹ (for forklaring av Mrk_dlx se ovenfor):</u>				
Mrk_dl1 = 1	-	-		
Mrk_dl1 = 2	-	-		
Mrk_dl1 = 3	-	-		
Mrk_dl1 = 4	-	-		
Mrk_dl1 = 6	-	-		
Mrk_dl1 = 7	-	-		
Mrk_dl1 = 8	-	-		
Mrk_dl1 = 9	9	1	2	1
Mrk_dl1 = A	-	-		
Mrk_dl1 = B	-	-		
Mrk_dl1 = Z	2171	2023	1925	1540
Mrk_dl2 = 0	233189	229652	228794	232223
Mrk_dl2 = 1	-	-		
Mrk_dl2 = 2	-	-		
Mrk_dl2 = 3	-	-		
Mrk_dl2 = 4	-	-		
Mrk_dl2 = 9	9	1	2	1
Mrk_dl3 = 0	231890	228156	227425	230982
Mrk_dl3 = 1	73	55	54	48
Mrk_dl3 = 2	1230	1435	1315	1192
Mrk_dl3 = 3	5	5	2	2
Mrk_dl3 = 4	-	-	-	-
Mrk_dl3 = 9	-	-	-	-
<u>Fødselsnummerkontroll:</u>				
Antall med gyldig fødselsnummer (Mrk_fnr = 0)	232664	228951	228823	231705
Antall med gyldig d-nummer (Mrk_fnr = 1)	-	-	-	-
Antall med nuller i personnummer (Mrk_fnr = 2)	-	-	-	-
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	534	702	513	519
Antall records i miniregisteret	233198	229653	228796	232224
Antall records på miniregisteret med utfall=8	207312	201698	201383	203989
<u>Utdanningsnivå (Miniregisteret):</u>				
Grunnskole				
Videregående skole				
Høgskole				
Universitet	23029	23493	22078	20761

¹ Start- og sluttdatoene for årgangene 1974/75 - 1985/86 er hentet fra variablene 'paabeg' og 'avslut'. Datomerkingen vil derfor avvike fra den vanlige for disse årgangene.

Avslutta	1979	1980	1981	1982
Antall records på grunnlagsregistrene	246754	255863	262081	273091
Antall records på grunnlagsregisteret med utfall=8	211477	226482	233264	245717
<u>Datomerking² (for forklaring av Mrk_dlx se ovenfor):</u>				
Mrk_dl1 = 1	-			
Mrk_dl1 = 2	-			
Mrk_dl1 = 3	-			
Mrk_dl1 = 4	-			
Mrk_dl1 = 6	-			
Mrk_dl1 = 7	-			
Mrk_dl1 = 8	-			
Mrk_dl1 = 9	-	6	5	4
Mrk_dl1 = A	-			
Mrk_dl1 = B	-			
Mrk_dl1 = Z	1693	1951	1701	1893
Mrk_dl2 = 0	244479	253531	259599	270989
Mrk_dl2 = 1	-			
Mrk_dl2 = 2	-			
Mrk_dl2 = 3	-			
Mrk_dl2 = 4	-			
Mrk_dl2 = 7	2			
Mrk_dl2 = 9	6	5	2	
Mrk_dl3 = 0	243213	252101	258322	269397
Mrk_dl3 = 1	51	50	48	71
Mrk_dl3 = 2	1214	1385	1234	1525
Mrk_dl3 = 3	1	1	-	-
Mrk_dl3 = 4	-	-	-	-
Mrk_dl3 = 9	-	-	-	-
<u>Fødselsnummerkontroll:</u>				
Antall med gyldig fødselsnummer (Mrk_fnr = 0)	243821	252953	259015	270547
Antall med gyldig d-nummer (Mrk_fnr = 1)	1	-	-	-
Antall med nuller i personnummer (Mrk_fnr = 2)	-	-	-	-
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	657	584	589	446
Antall records i miniregisteret	244479	253537	259604	270993
<u>Utdanningsnivå (Miniregisteret):</u>				
Grunnskole				
Videregående skole				
Høgskole				
Universitet	20998	21868	21790	24527

² Start- og sluttdatoene for årgangene 1974/75 - 1985/86 er hentet fra variablene 'paabeg' og 'avslut'. Datomerkingen vil derfor avvike fra den vanlige for disse årgangene.

Avslutta	1983	1984	1985	1986
Antall records på grunnlagsregistrene	302803	285184	295255	321060 ₃
Antall records på grunnlagsregisteret med utfall=8	250247	257655	268226	262379
<u>Datomerking⁴ (for forklaring av Mrk_dlx se ovenfor):</u>				
Mrk_dl1 = 1	-			
Mrk_dl1 = 2	-			
Mrk_dl1 = 3	-			
Mrk_dl1 = 4	-			
Mrk_dl1 = 6	-			
Mrk_dl1 = 7	-			
Mrk_dl1 = 8	-			
Mrk_dl1 = 9	5	15	94	2
Mrk_dl1 = A	-			
Mrk_dl1 = B	-			
Mrk_dl1 = Z	2987	2245	1880	1871
Mrk_dl2 = 0	276359	282903	293186	289831
Mrk_dl2 = 1	-			
Mrk_dl2 = 2	-			
Mrk_dl2 = 3	-			
Mrk_dl2 = 4	-			
Mrk_dl2 = 7	2	14	2	
Mrk_dl2 = 9	5	13	80	-
Mrk_dl3 = 0	274727	281125	291708	288249
Mrk_dl3 = 1	80	62	83	81
Mrk_dl3 = 2	1557	1731	1488	1503
Mrk_dl3 = 3	-	-	-	-
Mrk_dl3 = 4	-	-	-	-
Mrk_dl3 = 9	-	-	-	-
<u>Fødselsnummerkontroll:</u>				
Antall med gyldig fødselsnummer (Mrk_fnr = 0)	275908	282440	292698	288663
Antall med gyldig d-nummer (Mrk_fnr = 1)	-	1	-	-
Antall med nuller i personnummer (Mrk_fnr = 2)	1	-	-	-
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	455	477	582	1170
Antall records i miniregisteret	276364	282918	293280	289833
<u>Utdanningsnivå (Miniregisteret):</u>				
Grunnskole				
Videregående skole				
Høgskole				
Universitet	23975	23577	23760	22990

³ Dette antallet omfatter en del personer som har blankt utfall og avsluttadato utenfor perioden.

⁴ Start- og sluttdatoene for årgangene 1974/75 - 1985/86 er hentet fra variablene 'paabeg' og 'avslut'. Datomerkingen vil derfor avvike fra den vanlige for disse årgangene.

Avslutta	1987	1988	1989	1990
Antall records på grunnlagsregistrene	280624	284068	300076	331090
Antall records på grunnlagsregisteret med utfall=8	257601	259353	273415	298882
<u>Datomerking (for forklaring av Mrk_dlx se ovenfor):</u>				
Mrk_dl1 = 1	5797	74518	76720	75247
Mrk_dl1 = 2	7938	5165	5792	5640
Mrk_dl1 = 3	80665	17185	15945	17984
Mrk_dl1 = 4	167560	168190	174761	198495
Mrk_dl1 = 5	17584	17833	21173	26127
Mrk_dl1 = 6	-	1	-	-
Mrk_dl1 = 7	520	296	165	156
Mrk_dl1 = 8	-	-	4210	5221
Mrk_dl1 = 9	2	9	9	5
Mrk_dl1 = A	1	2	3	12
Mrk_dl1 = B	-	-	-	-
Mrk_dl1 = C	2264	2360	2647	2747
Mrk_dl1 = Z	-	-	-	-
Mrk_dl2 = 0	272749	276575	285368	311441
Mrk_dl2 = 1	9059	8677	11670	14799
Mrk_dl2 = 2	-	-	-	-
Mrk_dl2 = 3	-	-	-	-
Mrk_dl2 = 4	-	-	-	-
Mrk_dl2 = 7	276	169	12	322
Mrk_dl2 = 8	-	-	-	-
Mrk_dl2 = 9	247	138	4375	5072
Mrk_dl3 = 0	281637	285281	300653	330201
Mrk_dl3 = 1	7	15	59	99
Mrk_dl3 = 2	211	262	713	1333
Mrk_dl3 = 3	-	1	-	-
Mrk_dl3 = 4	-	-	-	1
Mrk_dl3 = 9	-	-	-	-
<u>Fødselsnummerkontroll:</u>				
Antall med gyldig fødselsnummer (Mrk_fnr = 0)	281637	284326	299861	329593
Antall med gyldig d-nummer (Mrk_fnr = 1)	-	-	-	-
Antall med nuller i personnummer (Mrk_fnr = 2)	-	2	-	1
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	694	1231	1564	2040
Antall records i miniregisteret	282331	285559	301425	331634
Antall records med utenlandsutdanning	2264	2360	2647	2747
<u>Utdanningsnivå (Miniregisteret):</u>				
Grunnskole				
Videregående skole				
Høgskole				
Universitet	17761	17963	21182	26137

Avslutta	1991	1992	1993	1994
Antall records på grunnlagsregistrene	337685	334901	357255	360201
Antall records på grunnlagsregisteret med utfall=8	306783	295936	312935	313490
<u>Datomerking (for forklaring av Mrk_dlx se ovenfor):</u>				
Mrk_dl1 = 1	75712	74997	74243	79101
Mrk_dl1 = 2	6312	6821	7002	6372
Mrk_dl1 = 3	13678	14769	14614	13627
Mrk_dl1 = 4	209904	199182	219376	218439
Mrk_dl1 = 5	29691	35770	38676	39830
Mrk_dl1 = 6	-	-	-	1
Mrk_dl1 = 7	623	145	77	48
Mrk_dl1 = 8	631	685	655	722
Mrk_dl1 = 9	2	2	3	19
Mrk_dl1 = A	25	97	1	5
Mrk_dl1 = B	-	-	-	-
Mrk_dl1 = C	2884	3104	2408	2422
Mrk_dl1 = Z				
Mrk_dl2 = 0	320713	314647	334954	338093
Mrk_dl2 = 1	17168	19996	21365	21699
Mrk_dl2 = 2	-	-	-	-
Mrk_dl2 = 3	-	-	-	-
Mrk_dl2 = 4	-	-	-	-
Mrk_dl2 = 7	602	223	215	276
Mrk_dl2 = 8	-	-	-	-
Mrk_dl2 = 9	667	706	521	518
Mrk_dl3 = 0	338295	334063	355221	358934
Mrk_dl3 = 1	64	110	209	183
Mrk_dl3 = 2	789	1383	1622	1464
Mrk_dl3 = 3	1	3	2	1
Mrk_dl3 = 4	1	13	1	4
Mrk_dl3 = 9	-	-	-	-
<u>Fødselsnummerkontroll:</u>				
Antall med gyldig fødselsnummer (Mrk_fnr = 0)	338390	334886	356184	359774
Antall med gyldig d-nummer (Mrk_fnr = 1)	-	-	-	-
Antall med nuller i personnummer (Mrk_fnr = 2)	1	10	5	8
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	759	676	866	804
Antall records i miniregisteret	339150	335572	357055	360586
Antall records med utenlandsutdanning	2884	3104	2408	2422
<u>Utdanningsnivå (Miniregisteret):</u>				
Grunnskole				
Videregående skole				
Høgskole				
Universitet	29707	35801	38775	39875
<u>Offisiell statistikk:</u>				
Grunnskole	
Videregående skole	200551	220432	221559	
Høgskole/Universitet	67774	70447	72960	

Avslutta	1995	1996	1997	1998
Antall records på grunnlagsregistrene	374804	395533	396650	398235
Antall records på grunnlagsregisteret med utfall=8	328973	351597	341862	341603
<u>Datomerking (for forklaring av Mrk_dlx se ovenfor):</u>				
Mrk_dl1 = 1	84577	92262	96269	95657
Mrk_dl1 = 2	3691	3121	1370	5500
Mrk_dl1 = 3	5537	3849	2521	1602
Mrk_dl1 = 4	233599	251005	240713	244405
Mrk_dl1 = 5	40954	42185	54160	46515
Mrk_dl1 = 6	-	-	1	-
Mrk_dl1 = 7	3206	442	5	419
Mrk_dl1 = 8	776	180	-	-
Mrk_dl1 = 9	-	3	2	1
Mrk_dl1 = A	49	6	6	7
Mrk_dl1 = B	-	-	-	-
Mrk_dl1 = C	2165	2527	2860	3196
Mrk_dl1 = Z				
Mrk_dl2 = 0	348405	372088	393501	372094
Mrk_dl2 = 1	22118	22861	4393	24781
Mrk_dl2 = 2	-	-	-	-
Mrk_dl2 = 3	-	-	-	-
Mrk_dl2 = 4	-	-	-	-
Mrk_dl2 = 7	2452	323	7	216
Mrk_dl2 = 8	-	-	-	-
Mrk_dl2 = 9	1579	308	6	211
Mrk_dl3 = 0	372902	394133	397272	394380
Mrk_dl3 = 1	201	201	104	207
Mrk_dl3 = 2	1443	1233	522	2712
Mrk_dl3 = 3	3	2	4	1
Mrk_dl3 = 4	5	11	5	2
Mrk_dl3 = 9	-	-	-	-
<u>Fødselsnummerkontroll:</u>				
Antall med gyldig fødselsnummer (Mrk_fnr = 0)	373540	394219	396665	395736
Antall med gyldig d-nummer (Mrk_fnr = 1)	1	6	3	8
Antall med nuller i personnummer (Mrk_fnr = 2)	8	8	1	3
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	1005	1347	1238	1555
Antall records i miniregisteret	372932	393719	395407	395120
Antall records med utenlandsutdanning	2165	2527	2860	3196
<u>Utdanningsnivå (Miniregisteret):</u>				
Grunnskole				
Videregående skole				
Høgskole				
Universitet	40981	42212	54343	47442
<u>Offisiell statistikk:</u>				
Grunnskole	55093	56026	53926	53056
Videregående skole	199159	243075	208421	214968
Høgskole/ Universitet	76046	96432	80487	73061

Avslutta	1999	2000	2001	2002
Antall records på grunnlagsregistrene	760717	807056		
Antall records på miniregisteret	593208	589248	597233	590863

Fødselsnummerkontroll:

Antall med gyldig fødselsnummer (Mrk_fnr = 0)	588087	582259	591259	582880
Antall med gyldig d-nummer (Mrk_fnr = 1)	1869	84	-	-
Antall med nuller i personnummer (Mrk_fnr = 2)	6	-	-	-
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	3246	6905	5974	7983

Antall records i miniregisteret	593208	589248	597233	590863
Avsluttede kurs	325353	318088	326076	313959
Eksamensrecorder (tilgangsdato er satt lik avgangsdato i NUDB)	267855	271160	271157	276904

Antall records med utenlandsutdanning	3359	3699	4969	4739
--	-------------	-------------	-------------	-------------

Utdanningsnivå (Miniregisteret, unntatt eksamensrecorder):

Grunnskolenivå	52144	52519	53445	56651
Videregående skole-nivå	230842	221262	226941	215103
Universitets- og høyskolenivå	42367	44307	45690	42205

Avslutta	2003	2004	2005
Antall records på grunnlagsregistrene			
Antall records på miniregisteret	768625	807006	817236

Fødselsnummerkontroll:

Antall med gyldig fødselsnummer (Mrk_fnr = 0)	758545	797954	
Antall med gyldig d-nummer (Mrk_fnr = 1)	-	1	
Antall med nuller i personnummer (Mrk_fnr = 2)	-	-	
Antall med ugyldig fødselsnummer (Mrk_fnr = 3)	10080	9051	

Antall records i miniregisteret	768625	807006	817236
Avsluttede kurs	328018	351596	356844
Eksamensrecorder (tilgangsdato er satt lik avgangsdato i NUDB)	440607	455410	460392

Antall records med utenlandsutdanning	4855	5252	6039
--	-------------	-------------	-------------

Utdanningsnivå (Miniregisteret, unntatt eksamensrecorder):

Grunnskolenivå	58022	62915	63795
Videregående skole-nivå	234349	250122	255209
Universitets- og høyskolenivå	35647	38559	37840

3.2.3. Kontroller på filer for befolkningens høyeste utdanning, per 1.okt. hvert år

Befolkningens høyeste utdanning	1970	1980	1985	1986
Antall records på grunnlagsregisteret	2 936 820	3 136 274	3 281 085	3 308 829
Antall dubletter på FNR (fjernet)	-	-	-	26
Antall dubletter på SNR (fjernet)	143	73	90	100
Antall personer på under 16 år i registeret	63 720	-	1 389	1 171
Antall personer med uoppgitt bosted i registeret	-	1 938	35	28
<u>Fødselsnummerkontroll i miniregisteret:</u>				
Antall med gyldig fødselsnummer	2 936 814	3 136 274	3 281 080	3 308 803
Antall med ugyldig fødselsnummer, slettes	6	-	5	-
<u>S-nummer kontroll i miniregisteret</u>				
Antall med ugyldig fnr (mrk_snr_nudb = '1')	-	-	-	-
Antall med 320's selvgenererte FNR, har 9 i syvende posisjon i FNR (mrk_snr_nudb = '2')	-	-	-	-
Antall gyldige fnr som ikke matcher mot SNR-katalogen (mrk_snr_nudb = '3') slettes	3	-	-	-
Antall records i miniregisteret	2 936 668	3 136 201	3 280 990	3 308 803
Antall i offisiell statistikk (Statistisk årbok)	2 873 114	3 134 528	3 279 661	3 307 630
<i>Avvik</i>	<i>63 554⁵</i>	<i>1 673</i>	<i>1 329</i>	<i>1 173</i>
<u>Datokontroll i miniregisteret (fullføringstidspunkt):</u>				
Antall records uten endring i fullføringstidspunkt (mrk_dl_bhu = '0')	-	3 040 960	3 190 873	3 213 157
Antall records som mangler fullføringstidspunkt; er satt til filas år og 06 (YYYY06) (mrk_dl_bhu = '1')	-	62 035	89 856	95 456
Antall records hvor månedsopplysningen er rettet fra '00' eller blank til '06' (mrk_dl = 2)	-	1	1	1
Antall records som er korrigert for ugyldig fullføringstidspunkt, (mrk_dl_bhu = '3')	-	33 278	355	189
Antall records hvor fullførtdato manglet på hele fila, og er satt til 197011 (mrk_dl_bhu = '4')	2 936 820	-	-	-
<u>Utdanningsnivå (miniregisteret⁶):</u>				
Grunnskole	-	-	-	1 201 413
Videregående skole	-	-	-	1 550 204
Universitets- og høyskole	-	-	-	460 909
Uoppgitt eller ingen fullført utdanning	-	-	-	96 277
<u>Offisiell statistikk:</u>				
Grunnskole	1 592 016 ⁷	1 349 382	1 228 874	1 201 407
Videregående skole	1 094 166	1 345 675	1 514 922	1 550 202
Universitets- og høyskole	230 051	378 056	446 627	460 909
Uoppgitt eller ingen fullført utdanning	20 435	63 088	90 567	96 185

⁵ Dette avviket ble nærmere beskrevet under kapitlet om 'Spesielle avvik i BHU-1970'.

⁶ Disse testene er gjort på grunnlag av den omkodede NUS89, da nivåinndelingen er endret noe i NUS2000-standarden og ikke er sammenliknbare med historiske tall fra før 1999.

⁷ Det er ikke publisert sammenliknbare omkodede tall for årene 1970,1980 og 1985.

Befolkningens høyeste utdanning	1987	1988	1989	1990
Antall records på grunnlagsregisteret	3 381 260	3 411 187	3 388 874	3 400 500
Antall dubletter på FNR (fjernet)	-	-	-	-
Antall dubletter på SNR (fjernet)	128	92	83	58
Antall personer under 16 år i registeret	1 064	989	978	818
Antall personer med uoppgitt bosted i registeret	25	24	18	20
<u>Fødselsnummerkontroll i miniregisteret:</u>				
Antall med gyldig fødselsnummer (mrk_fnr=0)	3 381 260	3 411 004	3 388 627	3 400 358
Antall med ugyldig fødselsnummer (mrk_fnr=3) slettes	-	183	247	142
<u>S-nummer kontroll i miniregisteret</u>				
Antall med ugyldig fnr (mrk_snr_nudb = '1')	-	-	-	-
Antall med 320's selvgenererte FNR, har 9 i syvende posisjon i FNR (mrk_snr_nudb = '2')	-	-	-	-
Antall gyldige fnr som ikke matcher mot SNR-katalogen (mrk_snr_nudb = '3')	-	-	-	-
Antall records i miniregisteret	3 381 132	3 410 912	3 388 544	3 400 300
Antall i offisiell statistikk (Statistisk årbok) ⁸	3 380 196	3 410 178	3 387 879	3 399 662
<i>Avvik</i>	936	734	665	638
<u>Datokontroll i miniregisteret (fullføringstidspunkt):</u>				
Antall records uten endring i fullføringstidspunkt (mrk_dl_bh = '0')	3 272 833	3 293 094	3 272 278	3 280 040
Antall records som mangler fullførings-tidspunkt; er satt til filas år og 06 (YYYY06) (mrk_dl_bh = '1')	108 388	118 091	116 594	88 932
Antall records hvor månedsopplysningen er rettet fra '00' eller blank til '06' (mrk_dl = 2)	1	1	1	31 527
Antall records som er korrigert for ugyldig fullføringstidspunkt, (mrk_dl_bh = '3')	38	1	1	1
Antall records hvor fullførtdato manglet på hele fila, og er satt til 197011 (mrk_dl_bh = '4')	-	-	-	-
<u>Utdanningsnivå (miniregisteret):</u>				
Grunnskole	1 199 265	1 172 893	1 122 921	1 096 003
Videregående skole	1 593 557	1 623 597	1 635 784	1 666 405
Universitets- og høyskole	479 235	495 812	512 814	548 427
Uoppgitt eller ingen fullført utdanning	109 075	118 610	117 025	89 465
<u>Offisiell statistikk:</u>				
Grunnskole	1 198 233	1 171 932	1 121 961	1 095 444
Videregående skole	1 593 558	1 623 954	1 635 783	1 666 405
Universitets- og høyskole	479 235	495 812	512 814	548 428
Uoppgitt eller ingen fullført utdanning	109 170	118 840	117 321	89 385

⁸ For 1987 er Statistisk Ukehefte nr. 36 1989 brukt som kilde, tall ikke tilgjengelig i Statistisk årbok for denne årgangen.

Befolkningens høyeste utdanning	1991	1992	1993	1994
Antall records på grunnlagsregisteret	3 416 459	3 433 126	3 448 098	3 461 479
Antall dubletter på FNR (fjernet)	-	-	-	-
Antall dubletter på SNR (fjernet)	55	53	51	27
Antall personer under 16 år i registeret	825	750	724	792
Antall personer med uoppgitt bosted i registeret	25	19	20	18
<u>Fødselsnummerkontroll i miniregisteret:</u>				
Antall med gyldig fødselsnummer (mrk_fnr=0)	3 416 304	3 433 073	3 447 923	3 461 275
Antall med ugyldig fødselsnummer (mrk_fnr=3) slettes	155	53	175	204
<u>S-nummer kontroll i miniregisteret</u>				
Antall med ugyldig fnr (mrk_snr_nudb = '1')	-	-	-	-
Antall med 320's selvgenererte FNR, har 9 i syvende posisjon i FNR (mrk_snr_nudb = '2')	-	-	-	-
Antall gyldige fnr som ikke matcher mot SNR-katalogen (mrk_snr_nudb = '3')	-	-	-	-
Antall records i miniregisteret	3 416 249	3 433 020	3 447 872	3 461 248
Antall i offisiell statistikk (Statistisk årbok)	3 415 609	3 432 357	3 447 354	3 460 669
<i>Avvik</i>	<i>640</i>	<i>663</i>	<i>518</i>	<i>579</i>
<u>Datokontroll i miniregisteret (fullføringstidspunkt):</u>				
Antall records uten endring i fullføringstidspunkt (mrk_dl_bhu = '0')	3 293 401	3 306 531	3 334 736	3 344 876
Antall records som mangler fullførings-tidspunkt; er satt til filas år og 06 (YYYY06) (mrk_dl_bhu = '1')	93 082	98 432	86 802	91 350
Antall records hvor månedsopplysningen er rettet fra '00' eller blank til '06' (mrk_dl = 2)	29 975	28 162	26 559	25 252
Antall records som er korrigert for ugyldig fullføringstidspunkt, (mrk_dl_bhu = '3')	1	1	1	1
Antall records hvor fullførtdato manglet på hele fila, og er satt til 197011 (mrk_dl_bhu = '4')	-	-	-	-
<u>Utdanningsnivå (miniregisteret):</u>				
Grunnskole	1 062 243	1 030 386	994 971	964 045
Videregående skole	1 688 377	1 706 710	1 729 257	1 743 554
Universitets- og høyskole	572 048	596 904	629 559	655 267
Uoppgitt eller ingen fullført utdanning	93 581	99 020	94 085	98 382
<u>Offisiell statistikk:</u>				
Grunnskole	1 061 452	1 029 666	994 257	963 265
Videregående skole	1 688 370	1 706 706	1 729 249	1 743 550
Universitets- og høyskole	572 049	596 903	629 558	655 266
Uoppgitt eller ingen fullført utdanning	93 738	99 082	94 290	98 588

Befolkningens høyeste utdanning	1995	1996	1997	1998
Antall records på grunnlagsregisteret	3 475 324	3 487 606	3 501 728	3 522 058
Antall dubletter på FNR (fjernet)	-	-	-	-
Antall dubletter på SNR (fjernet)	22	11	11	10
Antall personer under 16 år i registeret	793	811	805	788
Antall personer med uoppgitt bosted i registeret	18	17	14	13
<u>Fødselsnummerkontroll i miniregisteret:</u>				
Antall med gyldig fødselsnummer (mrk_fnr=0)	3 475 078	3 487 606	3 501 728	3 522 058
Antall med ugyldig fødselsnummer (mrk_fnr=3)	246	-	-	-
<u>S-nummer kontroll i miniregisteret</u>				
Antall med ugyldig fnr (mrk_snr_nudb = '1')	-	-	-	-
Antall med 320's selvgenererte FNR, har 9 i syvende posisjon i FNR (mrk_snr_nudb = '2')	-	-	-	-
Antall gyldige fnr som ikke matcher mot SNR-katalogen (mrk_snr_nudb = '3')	1	-	-	-
Antall records i miniregisteret	3 475 055	3 487 595	3 501 717	3 522 048
Antall i offisiell statistikk (Statistisk årbok)	3 474 513	3 486 778	3 500 909	3 521 257
<i>Avvik</i>	<i>542</i>	<i>817</i>	<i>808</i>	<i>791</i>
<u>Datokontroll i miniregisteret (fullføringstidspunkt):</u>				
Antall records uten endring i fullføringstidspunkt (mrk_dl_bh = '0')	3 356 717	3 368 289	3 378 051	3 390 180
Antall records som mangler fullførings-tidspunkt, er satt til filas år og 06 (YYYY06) (mrk_dl_bh = '1')	94 457	96 116	101 249	110 016
Antall records hvor månedsopplysningen er rettet fra '00' eller blank til '06' (mrk_dl=2)	24 149	23 200	22 427	21 861
Antall records som er korrigert for ugyldig fullføringstidspunkt, (mrk_dl_bh = '3')	1	1	1	1
Antall records hvor fullførtdato manglet på hele fila, og er satt til 197011 (mrk_dl_bh = '4')	-	-	-	-
<u>Utdanningsnivå (miniregisteret):</u>				
Grunnskole	908 844	856 523	824 719	795 564
Videregående skole	1 784 198	1 820 082	1 835 175	1 853 575
Universitets- og høyskole	680 839	708 164	734 142	756 711
Uoppgitt eller ingen fullført utdanning	101 174	102 826	107 681	116 198
<u>Offisiell statistikk:</u>				
Grunnskole	908 087	855 733	823 963	794 835
Videregående skole	1 784 193	1 820 076	1 835 169	1 853 569
Universitets- og høyskole	680 838	708 164	734 142	756 710
Uoppgitt eller ingen fullført utdanning	101 395	102 805	107 635	116 143

Befolkningens høyeste utdanning	1999	2000
Antall records på grunnlagsregisteret	3 541 202	3 559 199
Antall dubletter på FNR (fjernet)	-	-
Antall dubletter på SNR (fjernet)	7	-
Antall personer under 16 år i registeret	767	750
Antall personer med uoppgett bosted i registeret	11	11
<u>Fødselsnummerkontroll i miniregisteret:</u>		
Antall med gyldig fødselsnummer (mrk_fnr=0)	3 541 202	3 559 199
Antall med ugyldig fødselsnummer (mrk_fnr=3)	-	-
<u>S-nummer kontroll i miniregisteret</u>		
Antall med ugyldig fnr (mrk_snr_nudb = '1')	-	-
Antall med 320's selvgenererte FNR, har 9 i syvende posisjon i FNR (mrk_snr_nudb = '2')	-	-
Antall gyldige fnr som ikke matcher mot SNR-katalogen (mrk_snr_nudb = '3')	-	-
Antall records i miniregisteret	3 541 195	3 559 199
Antall i offisiell statistikk (Statistisk årbok)	3 540 424	
<i>Avvik</i>	<i>771</i>	
<u>Datokontroll i miniregisteret (fullføringstidspunkt):</u>		
Antall records uten endring i fullføringstidspunkt (mrk_dl_bhu = '0')	3 397 849	3 405 482
Antall records som mangler fullførings-tidspunkt, er satt til filas år og 06 (YYYY06) (mrk_dl_bhu = '1')	57 179	72 796
Antall records hvor månedsopplysningen er rettet fra '00' eller blank til '06' (mrk_dl = 2)	86 101	80 920
Antall records som er korrigert for ugyldig fullføringstidspunkt, (mrk_dl_bhu = '3')	73	1
Antall records hvor fullførtdato manglet på hele fila, og er satt til 197011 (mrk_dl_bhu = '4')	-	-
<u>Utdanningsnivå (miniregisteret):</u>		
Grunnskole	793 118	766 016
Videregående skole	1 942 526	1 956 018
Universitets- og høyskole	742 432	758 666
Uoppgett eller ingen fullført utdanning	63 119	78 499
<u>Offisiell statistikk⁹:</u>		
Grunnskole	792 437	
Videregående skole	1 942 521	
Universitets- og høyskole	742 423	
Uoppgett eller ingen fullført utdanning	63 034	

⁹ Disse testene er gjort på grunnlag av NUS2000, da disse filene kun inneholder 2000-standard og offisiell statistikk er beregnet ut i fra den. Nivå 5 i NUS2000 er inkludert i tall for videregående skole.

3.3. Avvik fra tidligere BHU-publisering

Etter at NUDB er etablert vil befolkningens høyeste utdanning(BHU) tas ut fra basen. Alle data, både fra igangværende, avslutta og bhu-filene danner grunnlaget for NUDB, og BHU vil derfor avvike fra tidligere kjøringer. Alle korreksjoner som er gjort på disse filene vil påvirke BHU-uttaket fra NUDB, bl.a. viser det seg at en del fullførte utdanninger fra avsluttafilene ikke er med på enkelte tidligere BHU-filer.

Opplysningene på BHU-miniregistre stemmer ikke med den offisielle statistikken som er publisert tidligere. Dette skyldes at en del personer under 16 år er inkludert på filene, samt at dubletter og records med ugyldige fødselsnummer er slettet. I grunnlagsregisteret er det også registrert noen med uoppgitt bosted, og disse blir ikke tatt med i den offisielle statistikken. Vi velger imidlertid å ta med disse gruppene inn i miniregisteret, og vil dermed kunne få avvik fra tidligere publisert statistikk.

Spesielle avvik for BHU-1970

BHU 1970 har større avvik enn de andre filene, dette skyldes av vi har foretatt et nytt uttrekk fra Folke- og bolig tellingen 1970 som benyttes i NUDB. I det nye uttrekket er alle over 16 år inkludert, i tillegg er personer over 13 år med avsluttet utdanning på utdanningsnivå mellom 1 og 8 også tatt med. Det ble også korrigeret for noen feil på klasstrinn: NUS73 4210 og klasstrinn '00', ble rettet til klasstrinn 12, og NUS73 4140 med klasstrinn '07' ble korrigeret til klasstrinn 12. I tillegg ble 14 dubletter slettet på grunnlag av laveste NUS og klasstrinn. Den nye BHU-1970 filen inneholder med andre ord 2.936.820 records, og ble kodet om til NUS89 og til slutt NUS2000.

I NUDB endres "Ingen utdanning" til "Uoppgitt utdanning" på dataene fra FOB70.










Begrunnelsen for å gjøre dette er følgende:

- 1) I Fob 70 er det registrert 20.000 med "Ingen utdanning" og 240 med "uoppgitt utdanning". Denne fordelingen mellom "ingen utdanning" og "uoppgitt utdanning" tyder på at gruppen som er kodet som ingen utdanning også er uoppgitt.
- 2) Folketellingsskjemaet for 1970 mangler rubrikk for "Ingen utdanning". Det vil si at den som ikke har utdanning bare skal unnlate å fylle ut spørsmålet.
- 3) I BHU 1980 er de som sto med "Ingen utdanning" i 1970 omkodet til "Uoppgitt". Det vil si at denne endringen allerede er gjort på tidligere BHU-filer.


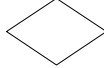
4. Diagrammer for miniregisterrutiner

Det er utarbeidet flere typer diagrammer over IT-rutinene. Diagrammene er også laget på flere nivåer. Dataflytdiagram nivå 1 gir en grov oversikt over systemet. Fysisk dataflytdiagram viser en detaljert oversikt over alle program, og sammenhengen mellom data og program. Det er utarbeidet diagrammer for rutiner i forbindelse med danning av miniregister i databasen. IT-dokumentasjonen er utarbeidet i henhold til SSBs egne metode for systemutvikling "Håndbok i utvikling av statistikkssystemer - Med vekt på IT-metode".

Standardsymboler fra IT-metoden som er benyttet i dataflytdiagrammene

Symbol	Symbolnavn	Tilhørende teknikk
	Database/datasett (ORACLE og SAS)	Fysisk dataflyt
	Sekvensielle filer (flate filer)	Fysisk dataflyt
	Dokument (papirtabeller)	Fysisk dataflyt
	Program	Fysisk dataflyt, dialogsystem, DFD0
	Dataflytretning	Fysisk dataflyt/ DFD0, DFD1, dialogsystem
	Peker til fortsettelse	Fysisk dataflyt
	Prosess	DFD1
	Datalager (entitet)	(DFD0), DFD1
	Fysisk datatabell i databasen	Datamodell

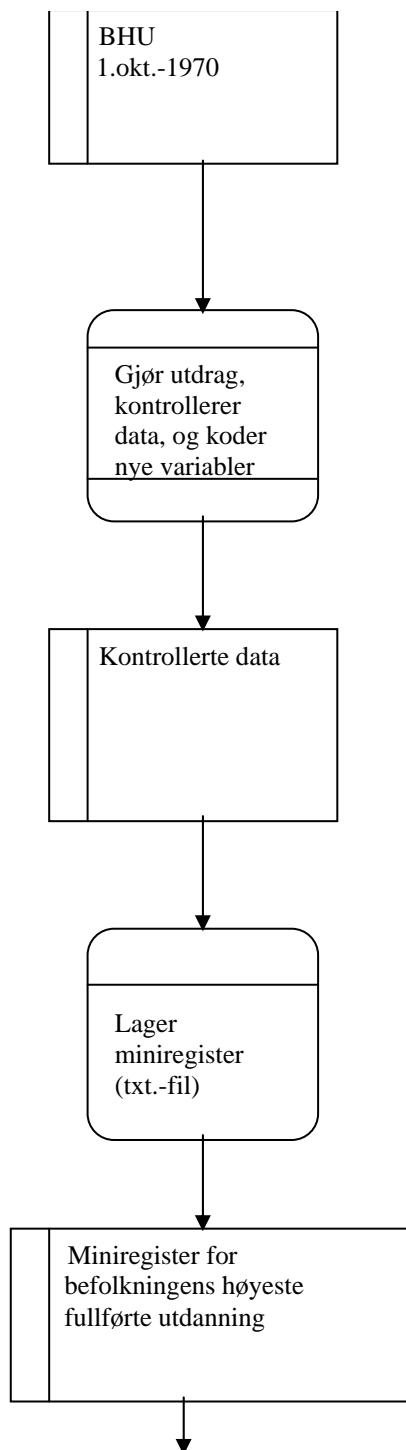
Standardsymboler i dataflytdiagram

Symbol	Symbolnavn	Tilhørende teknikk
	Arbeidsoperasjon	Flytdiagram for programrutiner
	Test	Flytdiagram for programrutiner

Dataflytdiagram nivå 1 - miniregister for BHU (SAS)

NUDB	Q:Dok\Utdstat\Nudb\Daytaflytdiagram\Bhu\Dataflyt_nivaa_1	
	Dataflytdiagramansvarlig: AsH	31/7-2001
Danning av miniregister for befolkningens høyeste fullførte utdanning (BHU)		

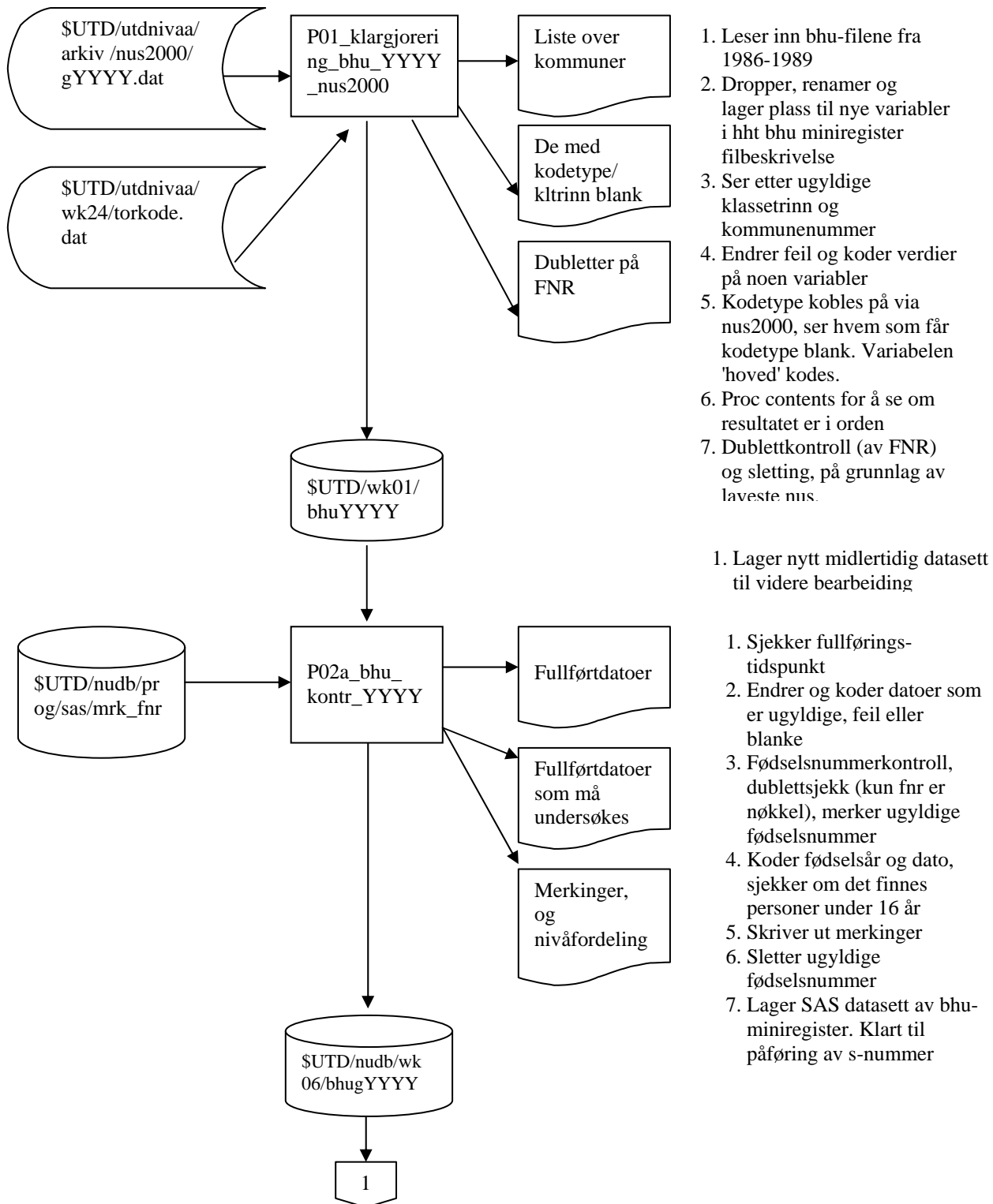
DFD1 gjelder tilsvarende for de senere årgangene



Til rutiner som danner forløp i NUDB

Fysisk dataflytdiagram - miniregister for BHU (SAS)

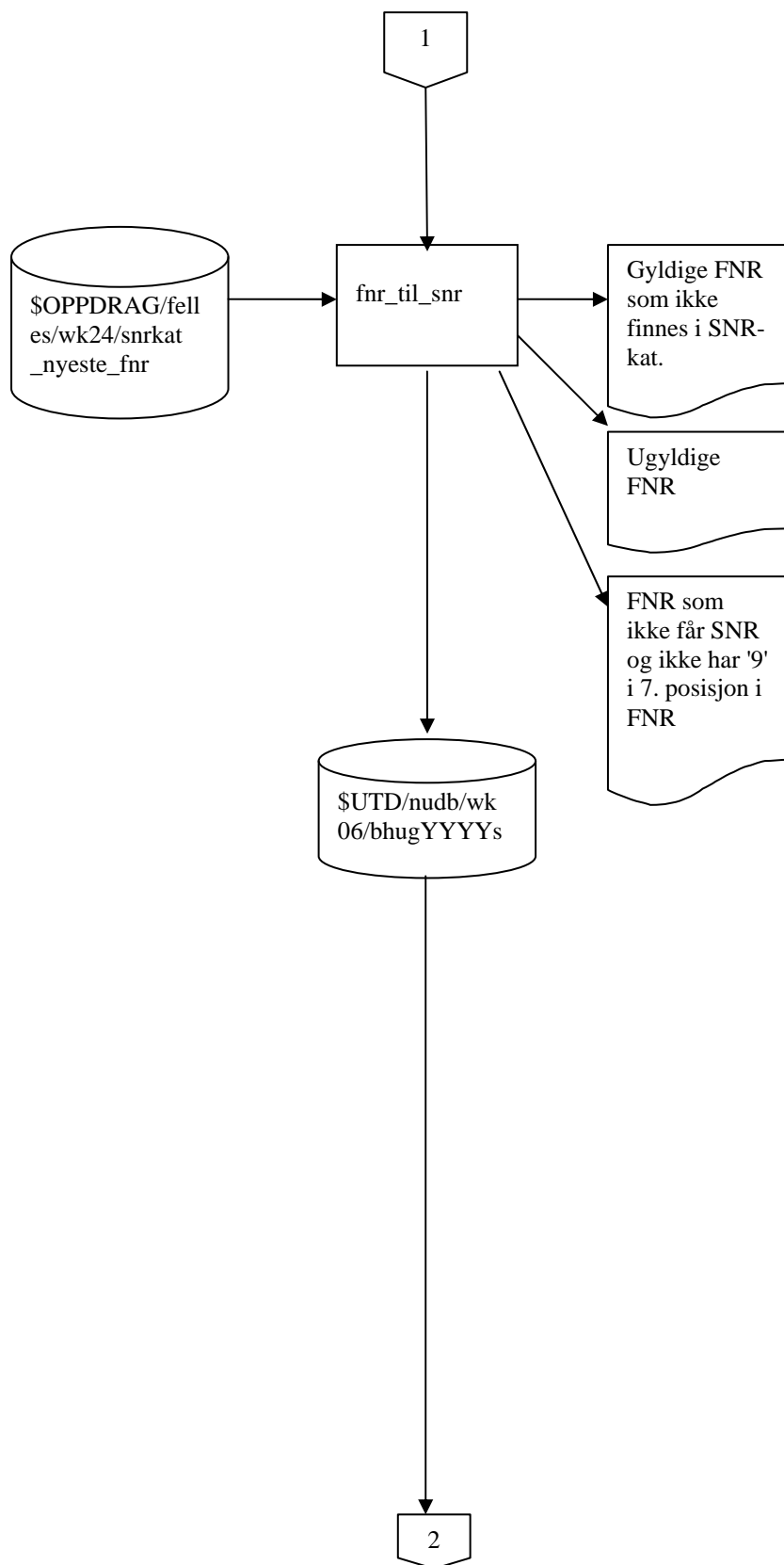
	Programmene ligger lagret under X:\360\Utd\NUDB\prog\bhu\miniregisterkontroll. Fnr_til_snr ligger lagret under: X:\320\1_ utdanning\utd\nudb\prog\felles\fnr_til_snr	
NUDB	Q:\Dok\Utdstat\Nudb\Daytaflytdiagram\Bhu\Dataflyt_bhu_1970_1985&1990_1999	
	Dataflytdiagramansvarlig: AsH	22/10-2001
Danning av miniregister for BHU 1970-1985 og 1990-1999		



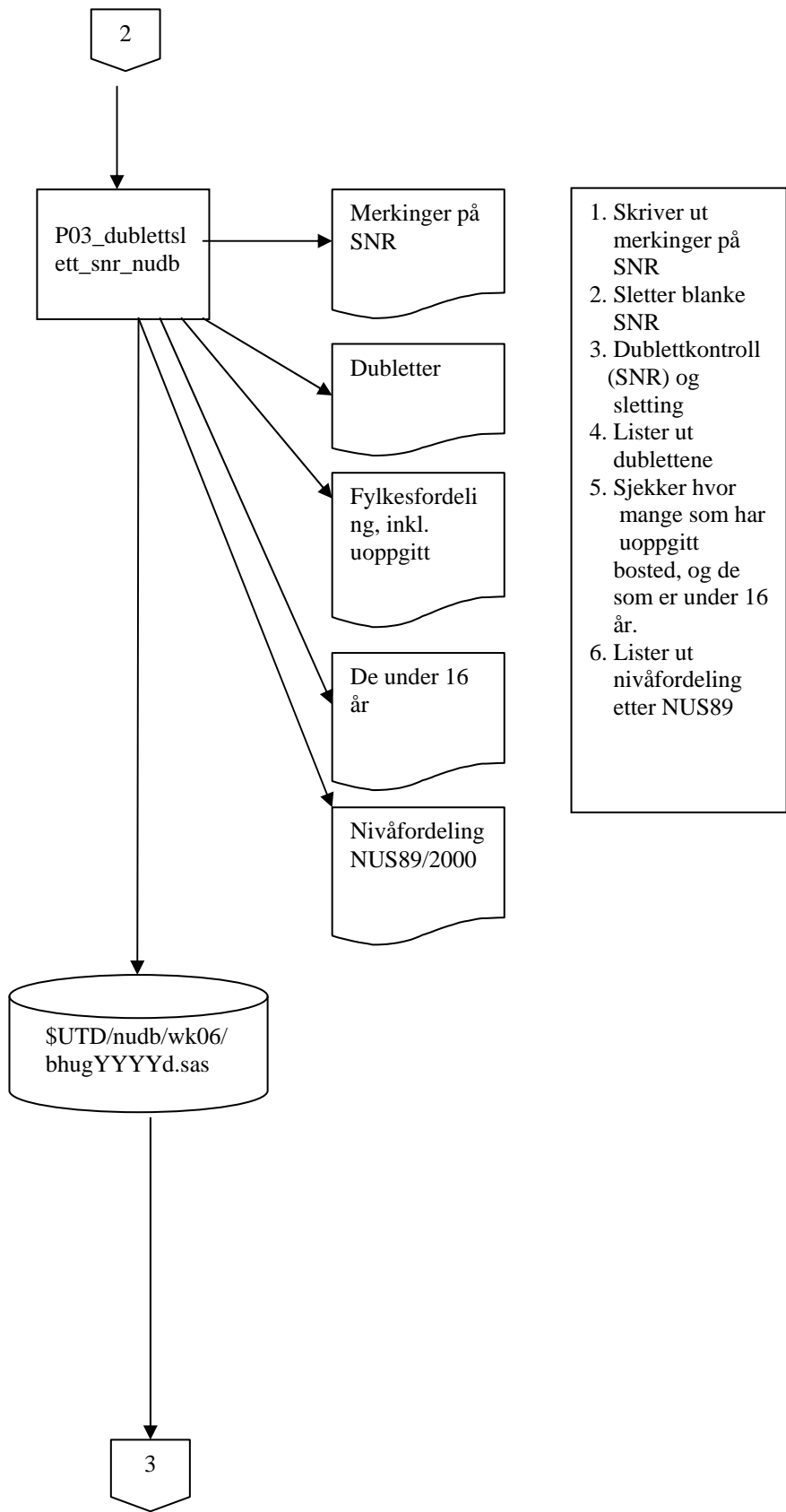
1. Leser inn bhu-filene fra 1986-1989
2. Dropper, renamer og lager plass til nye variabler i hht bhu miniregister filbeskrivelse
3. Ser etter ugyldige klassetrinn og kommunenummer
4. Endrer feil og koder verdier på noen variabler
5. Kodetype kobles på via nus2000, ser hvem som får kodetype blank. Variablen 'hoved' kodes.
6. Proc contents for å se om resultatet er i orden
7. Dublett kontroll (av FNR) og sletting, på grunnlag av laveste nus.

1. Lager nytt midlertidig datasett til videre bearbeiding

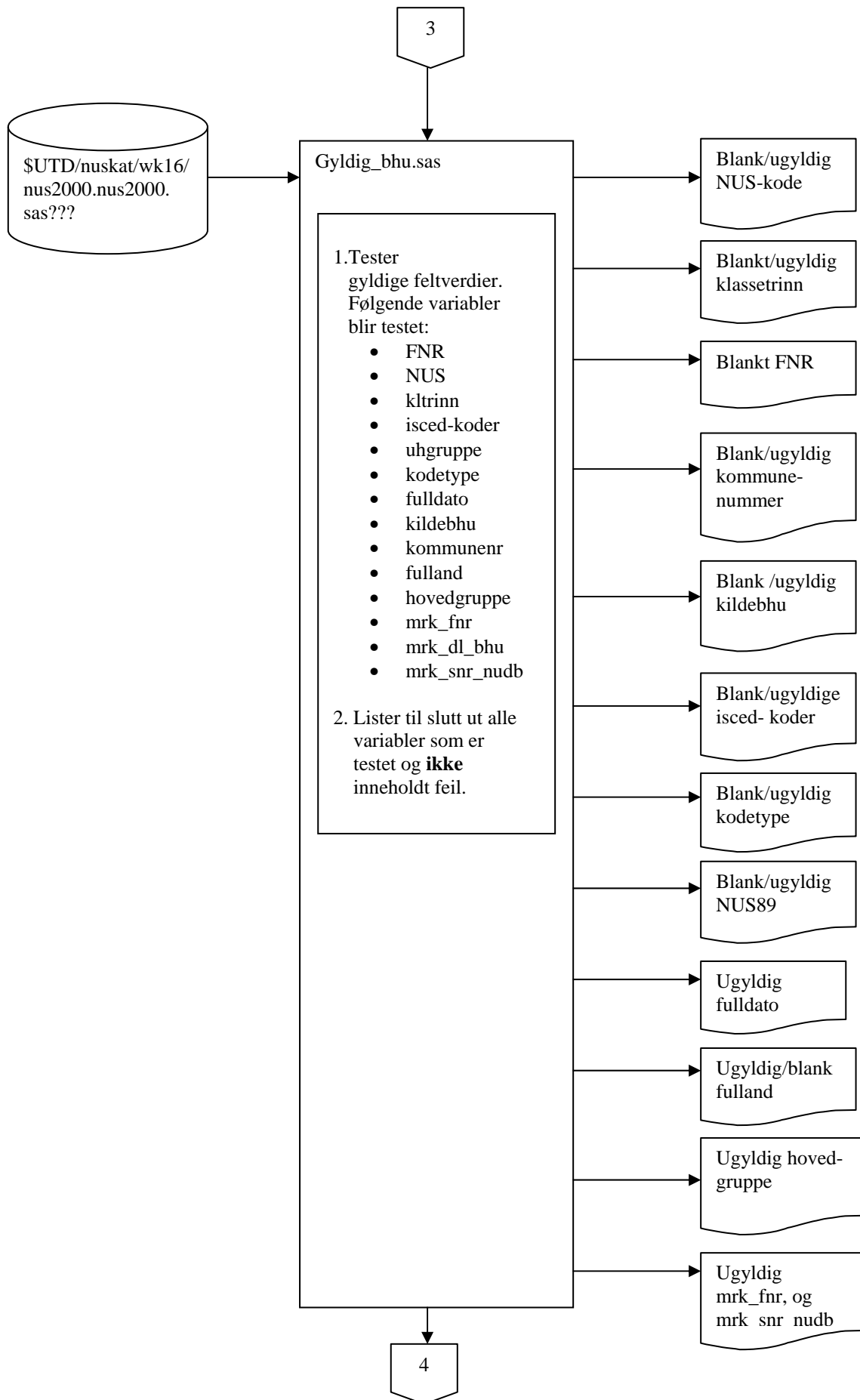
1. Sjekker fullførings-tidspunkt
2. Endrer og koder datoer som er ugyldige, feil eller blanke
3. Fødselsnummerkontroll, dublettsjekk (kun fnr er nøkkel), merker ugyldige fødselsnummer
4. Koder fødselsår og dato, sjekker om det finnes personer under 16 år
5. Skriver ut merkinger
6. Sletter ugyldige fødselsnummer
7. Lager SAS datasett av bhu-miniregister. Klart til påføring av s-nummer

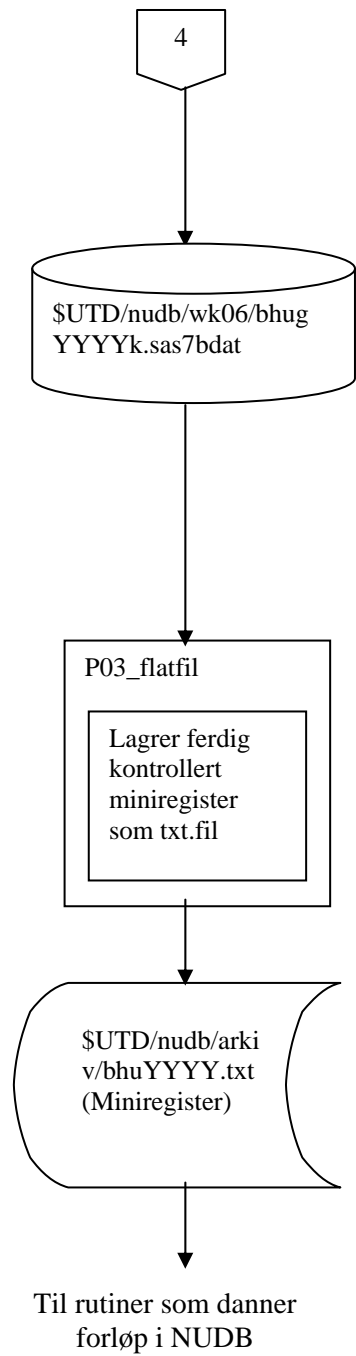


1. Omkoding fra FNR til SNR. Begge felt beholdes
2. Starter med en matematisk sjekk av FNR.
3. Merger filene med FNR som nøkkel
4. Merker ugyldige FNR, de får ikke påført SNR



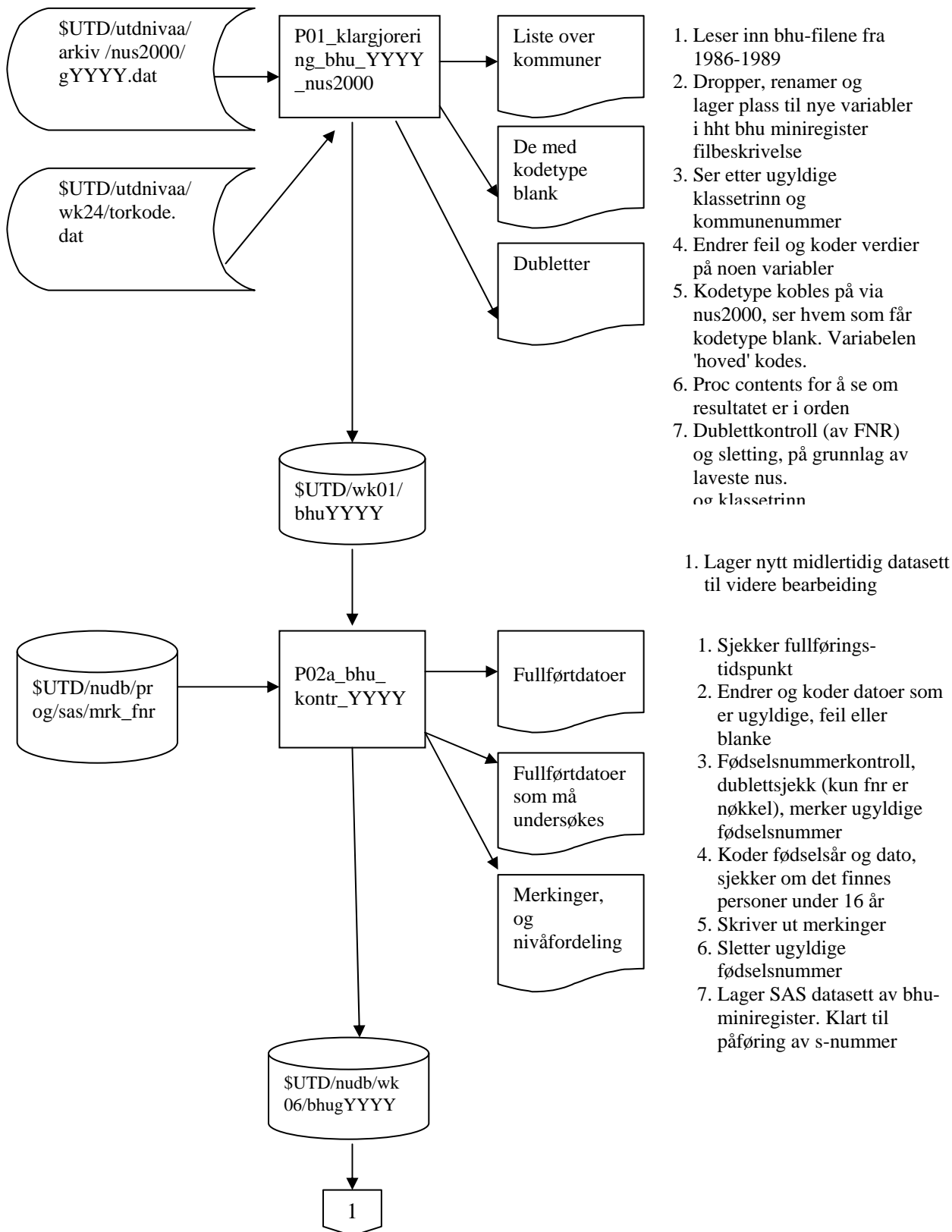
1. Skriver ut merkinger på SNR
2. Sletter blanke SNR
3. Dublett kontroll (SNR) og sletting
4. Lister ut dublettene
5. Sjekker hvor mange som har uoppgitt bosted, og de som er under 16 år.
6. Lister ut nivåfordeling etter NUS89

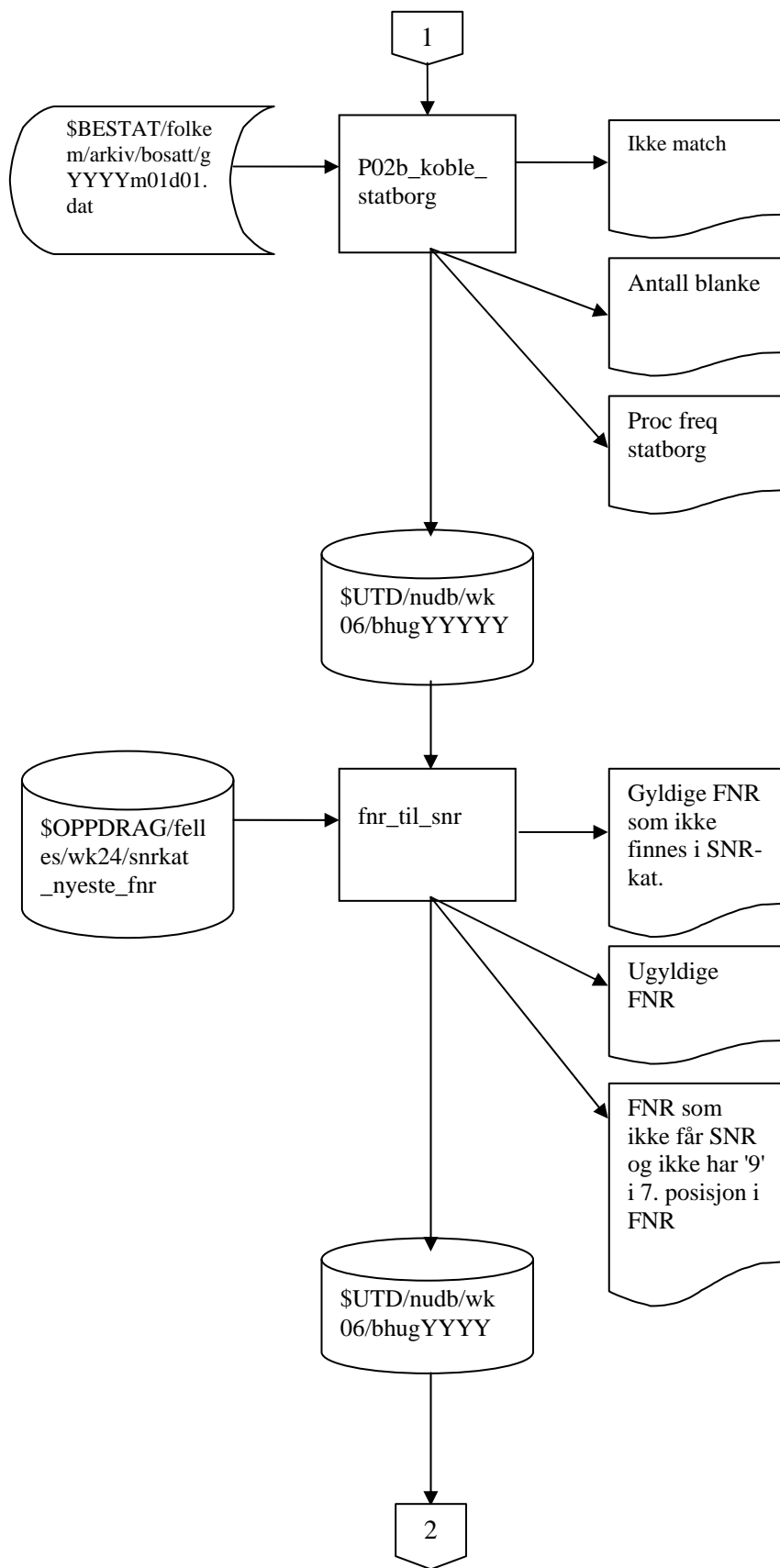




Fysisk dataflytdiagram - miniregister (SAS)

	Programmene ligger lagret under X:360\utd\nudb\prog\bhu\miniregisterkontroll. Fnr_til_snr ligger lagret under: X:360\utd\nudb\prog\felles\fnr_til_snr	
NUDB	Q:Dok\Utdstat\Nudb\Daytaflytdiagram\Bhu\Dataflyt_bhu_1986_1989	
	Dataflytdiagramansvarlig: AsH	31/7-2001
Danning av miniregister BHU 1986-1989		

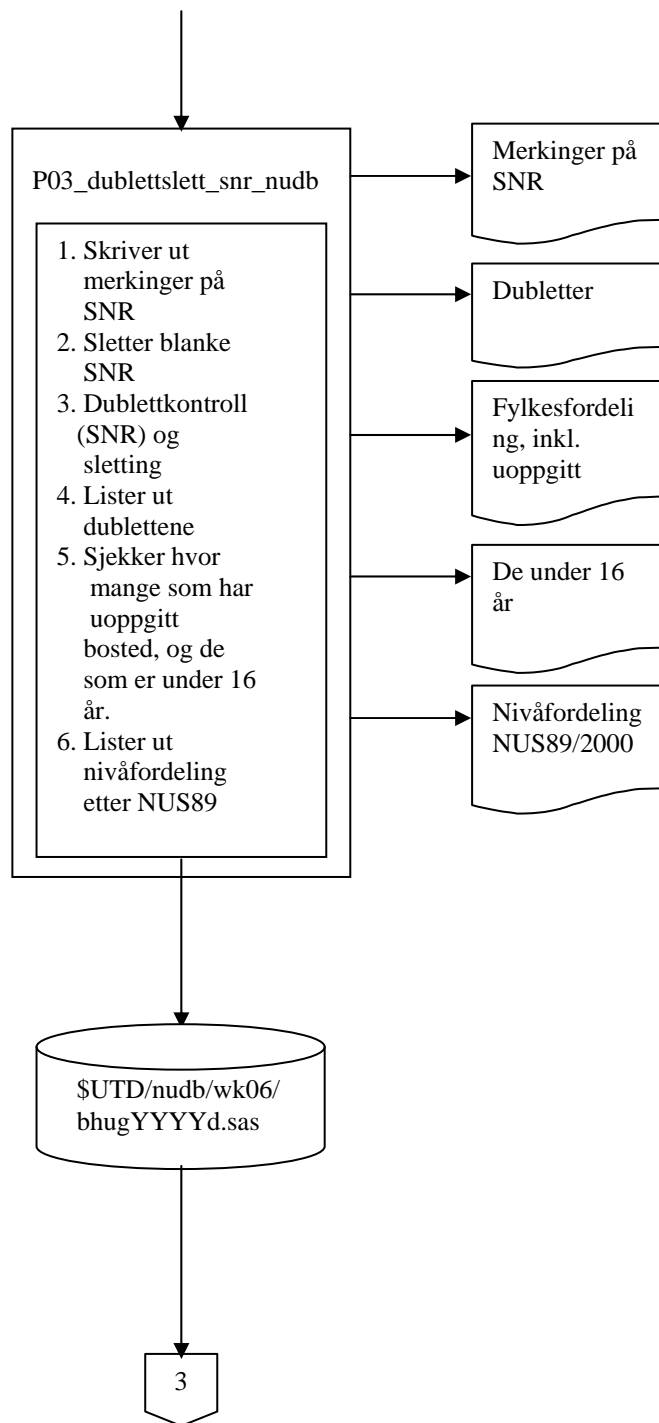


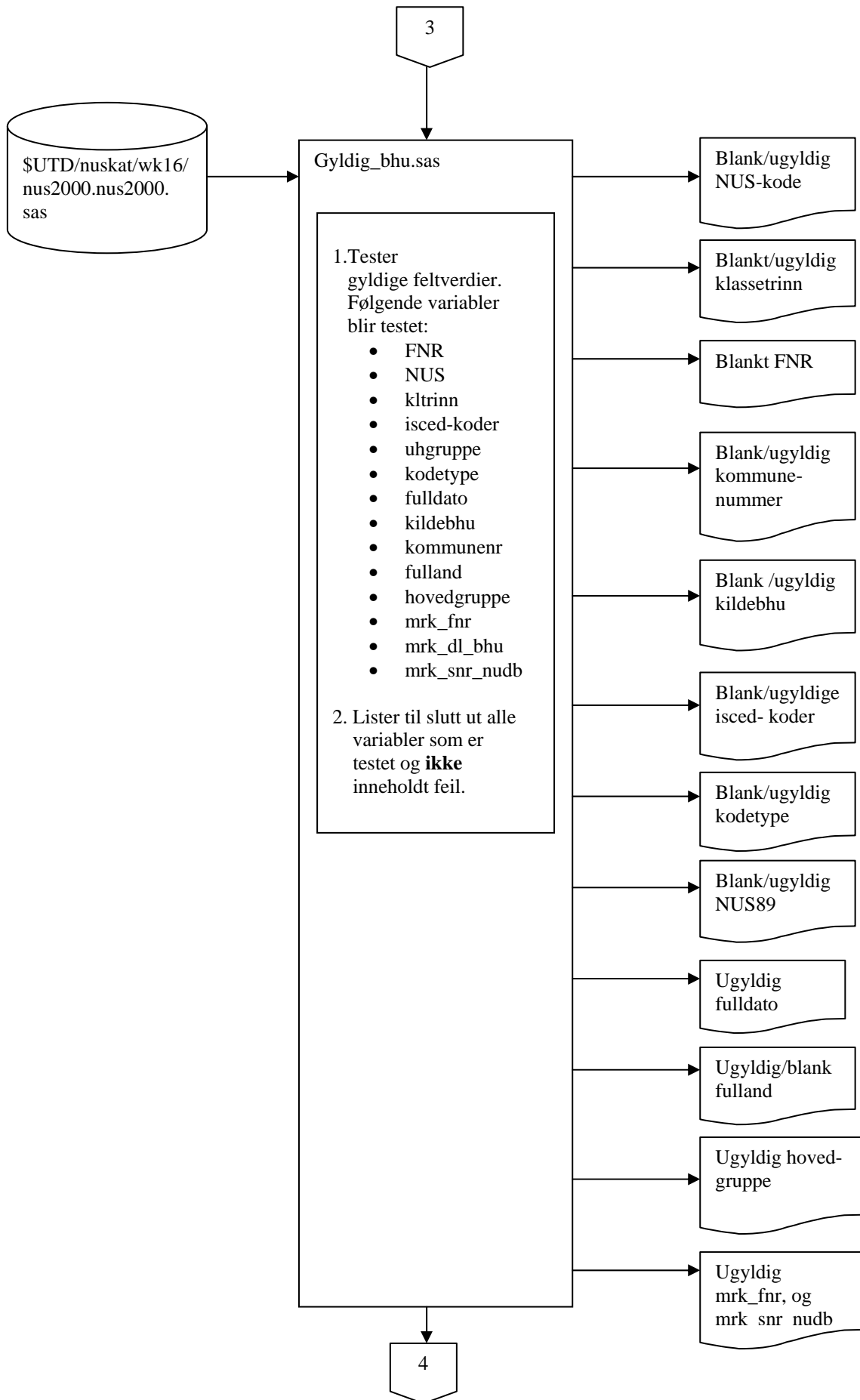


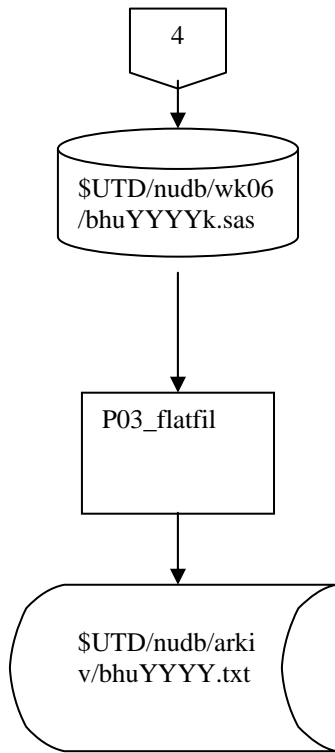
1. Kobler statsborgerskap på bhu-miniregister for årene 1986-1989
2. Merger statsborgerskap fra bosattfilen 1987-1990
3. Printer ut de som ikke får match
4. Setter de som ikke matcher eller er blank til '990' (uoppgitt)
5. Proc freq av resultat

1. Omkoding fra FNR til SNR. Begge felt beholdes
2. Starter med en matematisk sjekk av FNR.
3. Merger filene med FNR som nøkkel
4. Merker ugyldige FNR, de får ikke påført SNR

2





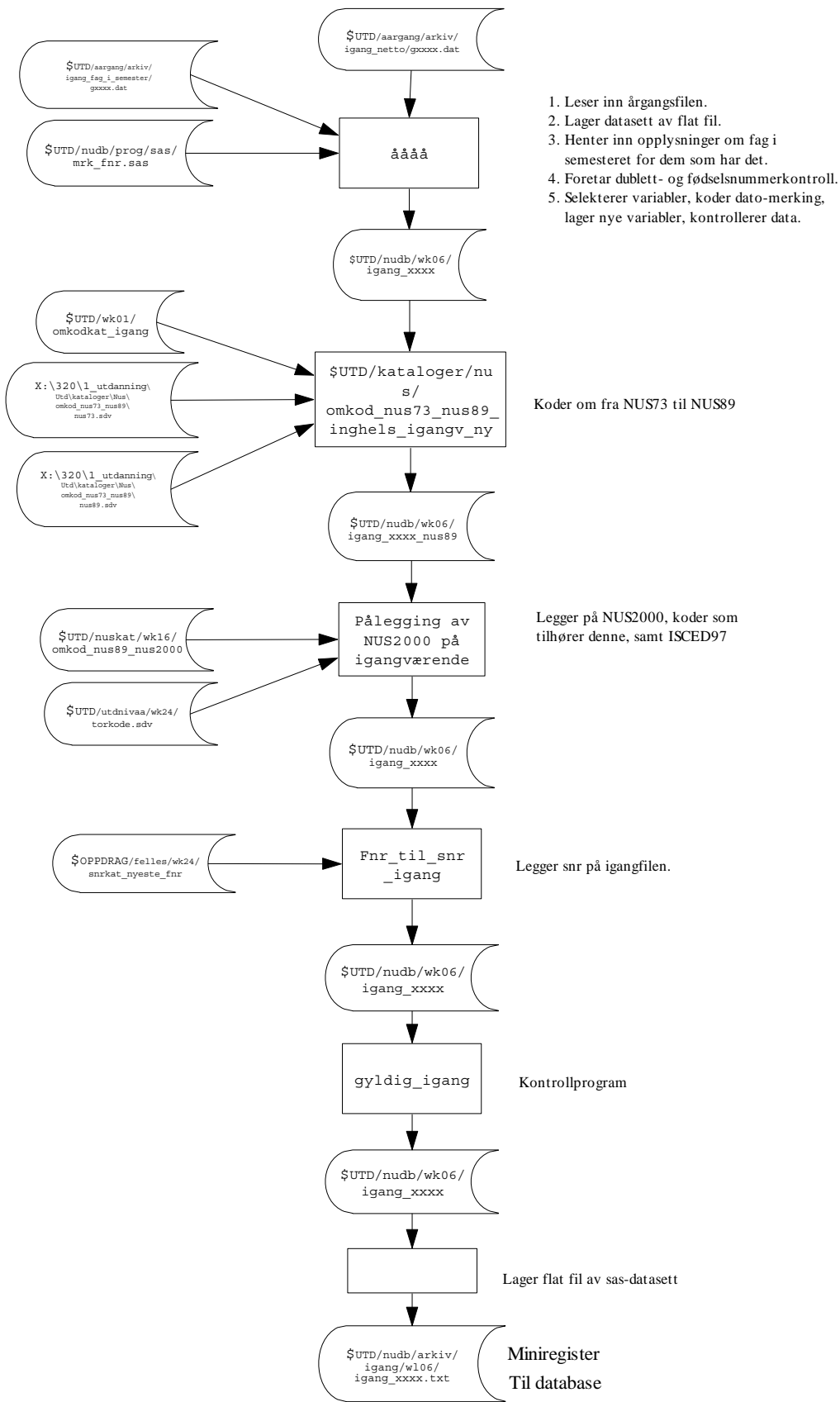


1. Lagrer ferdig miniregister som txt.fil

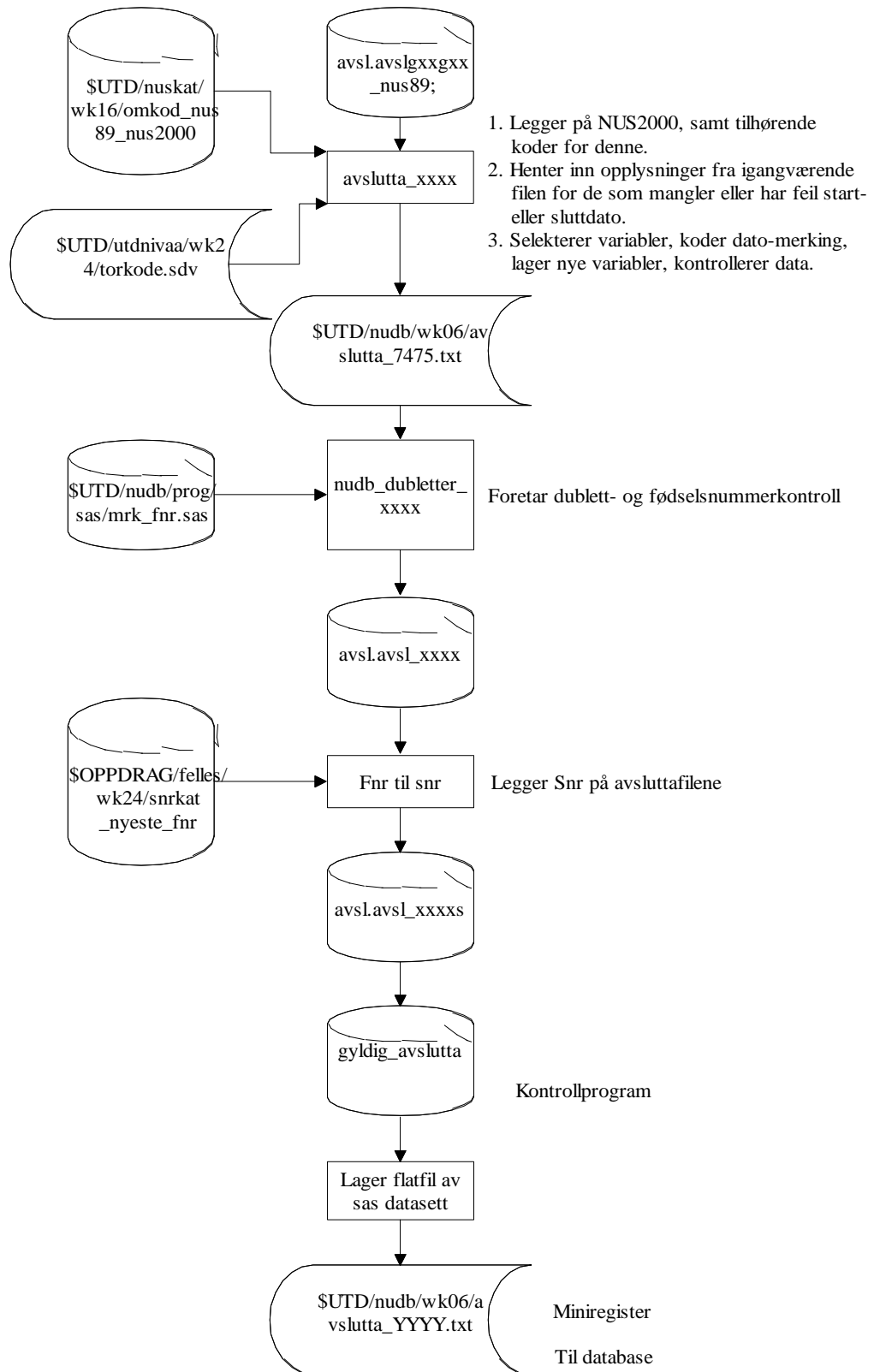
1. Miniregister

Til NUDB

Flytdiagram over omkoding av igangværende filer (1974-86) til miniregister i NUDB.
 Programmet ligger på X:\320\1_utdanning\Utd\nudb\prog\igangv\miniregisterprogrammer.
 Skrevet av Arild Rognan



Flytdiagram over omkodning av avsluttafiler (1974/75-1985/86) til miniregister i NUDB.
 Programmene ligger på X:\320\l_utdanning\Utd\nudb\prog\avslutt
 Skrevet av Arild Rognan



5. Tilrettelegging av forløp i NUDB

NUDB har tre forløpsfiler, en kursfil, en personfil med personens viktigste kurs og en demografifil.

5.1. Kursfilen. Forløp. (F_UTD_KURS)

Kursfilen er bygd opp av dataene fra de årlige filene for igangværende utdanning, avsluttet utdanning og de tidligere filene over befolkningens høyeste fullførte utdanning (BHU-filene). Fra og med 2001 bygger kursfilen i NUDB på de årlige filene over igangværende utdanning og avsluttet utdanning.

5.2. Viktigste kurs. Personnivået. Forløp. (F_UTD_PERSON)

5.2.1. Generelt

På personnivået skal alle de ulike kursene fra kursnivået knyttes sammen til utdanningstilfeller. Dette tilsvarer tabellen "viktigste kurs" i FD-Trygd. I Nudb har vi imidlertid behov for større detaljeringsgrad, og mer informasjon om utdanningene enn i FD-trygd-basen, og det må derfor utarbeides mer detaljerte regler.

Det vil være ulike definisjoner for hva som knyttes sammen til utdanningstilfeller på universitets- og høgskolenivået og på videregående nivået. Dette er fordi oppbyggingen av utdanninger på de to nivåene er veldig ulike, og skal tilfredstille ulike behov. Det godtas ikke overlappende utdanninger, så det er den til enhver tid viktigste utdanningen som velges ut fra bestemte kriterier og legges opp på personnivået. Det godtas heller ikke overlapping mellom nivåer, så her velges alltid utdanningen som er på det høyeste nivået.

Dersom en student er registrert med avlagte studiepoeng uten å være registrert som igangværende eller med fullført grad er denne ikke med på personnivået.

Det er den nye studiekoden som grupperer nuskoder sammen til utdanningstilfeller, se pkt 5.2.2.

På personnivået er det tre typer dateringer. Datering av en persons tilgang og avgang fra utdanning, datering av endret hovedgruppe (dvs. de tre gruppene grunnskole, videregående skole og høyere utdanning), og datering av endring i STUDIEKODE.

Definisjon av de tre avgangsdatoene som forekommer på personnivået:

AVGDATO: Tilfellets avgangsdato. Settes når en er ute av utdanningssystemet

UTDAVG: Avgangsdato for hovedgruppen. Hovedgruppe er 1, siffer i studiekode.

STUDAVG: Avgangsdato for studiet. Settes ved avgang fra en STUDIEKODE, eller ved overgang fra en studiekode til en ny.

Definisjon av de tre tilgangsdatoene som forekommer på personnivået:

TILGDATO: Tilfellets tilgangsdato. Settes når en person kommer inn i utdanningssystemet første gang eller etter å ha vært helt ute

UTDTILG: Tilgangsdato for hovedgruppen. Hovedgruppe er 1, siffer i studiekode.

STUDTILG: Tilgangsdato for studiet. Settes ved tilgang til en STUDIEKODE, eller ved overgang fra en studiekode til en ny.

Videregående nivå:

For videregående nivået er det den viktigste nuskoden som til enhver tid velges ut og legges opp på personnivået, sammen med utfall. Den nye studiekoden knytter nuskoder sammen til utdanningstilfeller. Den viktigste nuskoden (/utdanningen) velges etter bestemte regler. Se STUDIEKODE.

Det skal kodes variabler på normert tid og antall semester i den aktuelle utdanningen etter spesifiserte regler. På nivå 3, 4 og 5 er det kun utdanning under lov om videregående opplæring/opplæringsloven som det kodes normert tid for. se pkt 5.2.4.1.1.

Universitets- og høgsolenivået:

For universitets- og høgsolenivået vil vi også bruke den nye studiekoden (som for høyere utdanning bygger på UHGRUPPE-koden) for å knytte sammen nuskoder som hører sammen til et utdanningstilfelle. Studiekoden er laget for å gruppere NUS-koder som hører sammen. (Dette er de samme reglene som vil bli benyttet for koding av variablene for normert tid i høyere utdanning.)

Når en student tar flere utdanninger samtidig, som har ulike studiekoder velges den viktigste studiekoden etter bestemte regler Se pkt 5.2.3. Regelverk for valg av viktigste kurs.

Bare på det tidspunkt studenten avslutter en eksamen med nivå = 6, 7 eller 8, recordtype ulik 3 og utfall = 8 på kursnivået telles det opp hvor mange semester studenten har brukt på det aktuelle studiet, og om det er på normert tid se pkt. 5.2.4. Regler for telling av "antall semester innen studiet" ved fullføring av utdanning/grad.

5.2.2. STUDIEKODE på personnivået

STUDIEKODE	Tekst	Definisjon ¹⁾
100	Avsluttet grunnskole	MRK_GRSKOLE = 1
200	grunnkurs allmennfaglig studieretning	KTRINN=1 og STUDRETN=01,21-23 og HOVED=1,2
201	grunnkurs yrkesfaglig studieretning	KTRINN=1 og STUDRETN=03-19,31-99 og HOVED=1,2
210	vkI allmennfaglig studieretning	KTRINN=2 og STUDRETN=01,21-23 og HOVED=1,2
211	vkI yrkesfaglig studieretning	KTRINN=2 og STUDRETN=03-19,31-99 og HOVED=1,2
220	vkII allmennfaglig studieretning	KTRINN=3 og STUDRETN=01,21-23 og HOVED=1,2
221	vkII yrkesfaglig studieretning	KTRINN=3 og STUDRETN=03-19,31-99 og HOVED=1,2
230	andre kurs under lov om videregående opplæring	KTRINN=4 og HOVED=1,2
240	Fagopplæring før Reform94	KTRINN=blank og UTD=212 og HOVED=1,2
250	Andre videregående kurs ²⁾	KTRINN=blank og nivå=1-4 og HOVED=1,2 eller hvis KTRINN er utfylt og STUDRETN er blank og SUBSTR(NUS2000)=1-4
260	Andre kurs på mellomnivået ³⁾	nivå=5 og HOVED=1,2 eller hvis KTRINN er utfylt og STUDRETN er blank og SUBSTR(NUS2000)=5
301-399	Høyere utdanning	3+UHGRUPPE

¹⁾ Når et kurs innen videregående nivå, ifølge definisjonen kan havne i flere grupper, er det den laveste studiekoden som velges. ²⁾ Fagprøver fra 1988 og tidligere får studiekode 250 fordi de ikke kan skilles ut fra andre videregående uten å gå via teksten i NUS89. (Lærlinger er med i dataene kun fra og med 1989). ³⁾ Fra 2004/2005 vil fagskoleutdanning havne her fordi de nå har KTRINN=5. Tekniske fagskoleutdanning ble tidligere plassert under studiekode 230 fordi de da hadde KTRINN=4

Studiekoden for høyere utdanning har 3 i første posisjon, deretter følgende koder:

Nivå 6	Nivå 7	Nivå 8
<p>01 Forberedende prøver 02 Lavere nivå's utdanning 03 Andre ettårige studier, grunnutdanning 04 Høgskolekandidat, toårig 05 Ingeniørutdanning, toårig grunnutdanning 05H Høgskolekandidat, ingeniørfag, toårig 06 Andre toårige studier, grunnutdanninger 07 Høgskolekandidat, treårig 08 Allmennlærer, grunnutdanning 08B Bachelor, allmennlærerutdanning (foreløpig bare ved Rudolf Steinerhøgskolen) 09 Førskolelærer, grunnutdanning 09B Bachelor, førskolelærerutdanning 10 Fag- og yrkesfaglærer, grunnutdanning 10B Bachelor, fag- og yrkesfaglærerutdanning 11 Ingeniørutdanning, treårig grunnutdanning 11B Bachelor, ingeniørfag 12 Sykepleieutdanning, grunnutdanning 12B Bachelor, sykepleieutdanning 13 Helsefagutdanning, 3-4 år grunnutdanning, ikke sykepleier 13B Bachelor, helsefagutdanning, ikke sykepleier 14 Høgskolekandidat, fireårig 14B Bachelor, fireårig 15 Etatsutdanninger 16 Andre tre- og fireårige grunnutdanninger (ikke høgskolekandidat) 17 Cand.mag.-utdanning 18 Videreutdanninger i ledelse/organisasjon/administrasjon/økonomi 19 Videreutdanning for ingeniører 20 Videreutdanning for sykepleiere 21 Videreutdanning for annet helsepersonell enn sykepleiere 22 Annen videreutdanning, til og med to år 23 Praktisk-pedagogisk utdanning (PPU) 24 Siviløkonomutdanning, fireårig 25B Bachelor, allmenne fag 26B Bachelor, humanistiske og estetiske fag 27B Bachelor, lærerutdanninger og utdanninger i pedagogikk, ikke allmennlærer 28B Bachelor, samfunnsfag og juridiske fag 29B Bachelor, økonomiske og administrative fag 30B Bachelor, naturvitenskapelige fag, håndverksfag og tekniske fag 96B Bachelor, helse-, sosial- og idrettsfag 97B Bachelor, primærnæringsfag 98B Bachelor, samferdsels- og sikkerhetsfag og andre servicefag 99B Bachelor, uoppgitt fagfelt 69M Master, ettårig, lavere nivå</p>	<p>31 Høyere nivå's utdanning 32 Cand.philol.-utdanning 33 Cand.polit.-utdanning 34 Cand.scient./Cand.real.-utdanning 35 Cand.jur.-utdanning 35M Master, rettsvitenskap 36 Cand.med.-utdanning 37 Cand.agric.-utdanning 38 Cand.musicae.-utdanning 39 Cand.theol.-utdanning 40 Cand.san.-utdanning 41 Cand.oecon.-utdanning 42 Cand.psychol.-utdanning 43 Cand.sociol.-utdanning 44 Cand.sosion.-utdanning 45 Cand.act.-utdanning 46 Cand.techn.-utdanning 47 Cand.pharm.-utdanning 47M Master, farmasi-utdanning 48 Cand.med.vet.-utdanning 49 Cand.merc.-utdanning 50 Cand.odont.-utdanning 50M Master, odontologi-utdanning 51 Cand.ed.-/cand.paed.-/cand.paed.spec.-utdanning. 52 Hovedfagskandidat 53 Magisterutdanning 54 Siviløkonomutdanning, CEMS-master 54M Siviløkonom-/Master-utdanning 55 Sivilingeniørutdanning 55M Master, teknologifag 56 Master of Science 57 Master of Philosophy 58 Master of Arts 59 Master of International Business 60 Master of Business Administration 61 Master of Management 62 Master of Technology Management 63 Andre mastergradsutdanninger 64 Master of Laws 65 Annen utdanning på nivå 7, ikke videreutdanning/påbygging 66 Videreutdanninger/påbygging, inntil to år 67 Spesialistutdanning for leger 68 Videreutdanning for sykepleiere på nivå 7 71M Master, allmenne fag 72M Master, humanistiske og estetiske fag 73M Master, lærerutdanninger og utdanninger i pedagogikk 74M Master, samfunnsfag og juridiske fag 75M Master, økonomiske og administrative fag 76M Master, naturvitenskapelige fag, håndverksfag og tekniske fag 77M Master, helse-, sosial- og idrettsfag 78M Master, primærnæringsfag 79M Master, samferdsels- og sikkerhetsfag og andre servicefag 80M Master, uoppgitt fagfelt</p>	<p>81 Dr.philos. 82 Dr.polit. 83 Dr.scient. 84 Dr.juris./dr.legis. 85 Dr.med. 86 Dr.theol. 87 Dr.techn. 88 Dr.odont. 89 Dr.med.vet. 90 Dr.agric. 91 Dr.oecon. 92 Dr.ing. 93 Dr.art. 94 Dr.psychol. 95 Andre forskerutdanninger (lisensiat)er 70 Ph.d.</p>

5.2.3. Regelverk for valg av viktigste kurs.

Prioriteringsrekkefølge

0. Utenlandsstudenter (SKOLEKOM='2580') velges alltid sist

1. Høyere nivå (HOVED=3) velges foran lavere nivå (HOVED=1-2)

Innen høyere nivå (HOVED=3) velges det slik:

- Etter nivå (første siffer i nus2000): nivå 8 velges før nivå 7, som igjen velges før nivå 6.
- etter STUDIEKODE med følgende prioriteringer: alle verdier er viktigere enn 01, 02 og 03
- Deretter velges den som har fullført før den som ikke er fullført.
- Utdanning ved vitenskapelige høyskoler beholdes fremfor all annen utdanning. Dette selekteres etter organisasjonsnummer. Følgende er de vitenskapelige høyskolene, etter org.nummer (inr på tidligere åranganger):
 - 974705230 (0993131) Norges landbrukshøgskole
 - 974799111 (1965719) Norges handelshøgskole
 - 974714442 (1965670) Norges veterinærhøgskole
 - 971526033 (1965700) Norges idrettshøgskole
 - 971526378 (1965697) Arkitektshøgskolen i Oslo
 - 970011994 (1965689) Menighetsfakultetet
 - 874714712 (1958453) Norges Musikkhøgskole
 - 976776070 (7963998) Norges musikkhøgskole, studiested Carl Berner
- Lærerutdanning beholdes før andre utdanninger/grunnfag. Dette selekteres på studiekodene 308, 309 og 310.
- På nivå 8 (studiekode 381-395, 370): alle andre studiekoder beholdes før studiekode 395. På nivå 7 (studiekode 331-367,371-380): studiekode 332-363, 371-379 beholdes før studiekode 331, 364-367, 380.
- På nivå 6 (studiekode 301-330, 396-399): skal de gjenværende dublettene selekteres etter studiekode i følgende rekkefølge (de som kommer først i listen velges foran de som kommer senere i lista):

studiekode 324	siviløkonom
studiekode 308, 314, 317	fireårig
studiekode 313, 316	3-4 år
studiekode 307, 309-312	treårig
studiekode 304-306	toårig
studiekode 325-330, 396-398	bachelor
studiekode 315, 318-323	etatsutd + videreutdanning
studiekode 302, 303	ettårig
studiekode 301	forberedende prøver
- Dersom det fremdeles finnes dubletter igjen innen hoved=3 velges tilfeldig(den høyeste verdien).

2. Innen lavere nivå velges HOVED=2 foran HOVED=1.

3. Innen HOVED=1 velges tilfeldig.

4. Innen HOVED=2 velges

- kurs med KOMP ulik blank framfor de med KOMP = blank.
- Blant de med KOMP = blank velges kurs med KTRINN utfylt framfor KTRINN = blank.
- Blant de med KOMP ulik blank velges tilfeldig
- Blant de med KTRINN utfylt velges KTRINN=3 foran KTRINN=2 foran KTRINN=1 foran KTRINN=4. - Blant de med KTRINN = blank velges den med høyeste klasstrinn.
- Hvis klasstrinn er like velges tilfeldig.

5.2.4. Regler for telling av "antall semester innen studiet" ved fullføring av utdanning/grad

Variabelen SEMESTER_INNEN_STUDIE beregnes for hver gang en student fullfører treårig videregående eller en grad i høyere utdanning. Telling av semester ved fullført utdanning gjøres bare der det er mulig å definere hvilke utdanninger som inngår i en grad. Denne tellingen av brukte semestre benyttes også for å sette variabelen "Fullført på normert tid"

Opptelling av semester innen utdanningen for videregående gjøres når et kurs med utfall 8 og
- KOMP=1,2,3, 4, 5 forekommer eller
- KOMP=blank og KTRINN er utfylt og KLASSETRINN =13 og UTD ulik 212 (vk2 unntatt fagprøve) eller
- KOMP=blank og KTRINN er utfylt og KLASSETRINN =14 og UTD ulik 212 (vk3 unntatt fagprøve).

5.2.5. Regler for å telle normert tid for de som har fullført en utdanning

5.2.5.1. Videregående utdanning:

For videregående utdanning (nivå 3 og 4) er det tatt utgangspunkt i Kompetansekoden

Hvis kompetansekode er utfylt så settes normert varighet slik:

Kompetanse-kode		Studiekoder som inngår ved telling av semester	Normert tid, antall sem.
1	VK II gir generell studiekompetanse	200-230	6
2	VKII gir yrkeskompetanse med svennebrev eller fagbrev og VK II foregår i bedrift	200-230	8
3			
4	VK II gir yrkeskompetanse uten svennebrev eller fagbrev	200-230	6
5	VK II gir yrkeskompetanse med svennebrev eller fagbrev, men VK II foregår i skole før læretid i bedrift.	200-230	6
	VK III gir yrkeskompetanse med svennebrev eller fagbrev, men VK II foregår i skole før læretid i bedrift.	200-230	10

Ellers:

		Studiekoder som inngår ved telling av semester	Normert tid, antall sem.
Hvis kompetansekode(KOMP) = blank og kurstrinn(KTRINN) er utfylt og klassetrinn(KLTRINN2000) = 13 og UTD ulik 212*	vkII under lov om vid.opplæring/opplæringsl oven	200-230	6
Hvis kompetansekode = blank og kurstrinn er utfylt og klassetrinn = 14 og UTD ulik 212	vkIII under lov om vid.opplæring/opplæringsl oven	200-230	8

Ellers

skal det ikke settes noen verdi for normert tid.

5.2.5.2. Høyere utdanning:

Det telles normert tid kun for de som har fullført (og bestått) en grad (utdanning med recordtype ulik 3).

Studiekode brukes for å bestemme hvilke utdanninger/kurs/NUS-koder som hører sammen. Når en utdanning innen en studiekode er fullført, telles antall semester bakover i tid innen de studiekodene som er definert nedenfor (à jour per 2005). Om utdanningen er fullført på normert tid bestemmes mot

en katalog der normert tid for studiet er angitt. Denne katalogen er ikke publisert her, men er lagret i NUDB.

Avsl. utd. innen studiekode:

Gyldige kombinasjoner av studiekode:

301 Forberedende prøver	*
302 Lavere nivå's utdanning	*
303 Andre ettårige studier, grunnutdanning	*
304 Høgskolekandidat, toårig	304
305 Ingeniørutdanning, toårig grunnutdanning	305
306 Andre toårige studier, grunnutdanninger	306
307 Høgskolekandidat, treårig	307,
308 Allmennlærer, grunnutdanning	308, 302, 303, 322
308B Bachelor, allmennlærerutdanning	308, 308B
309 Førskolelærer, grunnutdanning	309, 322
309B Bachelor, førskolelærerutdanning	309B, 309, 322
310 Yrkesfaglærer, grunnutdanning	*
310B Bachelor, yrkesfaglærerutdanning	310B,310
311 Ingeniørutdanning, treårig grunnutdanning	311,
311B Bachelor, ingeniørfag	311B, 311
312 Sykepleierutdanning, grunnutdanning	312
312B Batchelor, sykepleieutdanning	312B, 312
313 Helsefagutdanning, 3-4 år grunnutdanning, ikke sykepleier	313
313B Batchelor, helsefagutdanning, ikke sykepleier	313B, 313
314 Høgskolekandidat, fireårig	314, 304, 307
314B Høgskolekandidat, fireårig	314, 304, 307, 314B
315 Etatsutdanninger	*
316 Andre tre- og fireårige grunnutdanninger (ikke høgskolekandidat)	316, 302, 303, , 306,
317 Cand.mag.-utdanning	317, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 316, 318, 319, 320, 321, 322, 323
318 Videreutdanninger i ledelse/organisasjon/administrasjon/økonomi	*
319 Videreutdanning for ingeniører	*
320 Videreutdanning for sykepleiere	*
321 Videreutdanning for annet helsepersonell enn sykepleiere	*
322 Annen videreutdanning, til og med to år	*
323 Praktisk-pedagogisk utdanning (PPU)	*
324 Siviløkonomutdanning	324, 304, 306, 316
325B Bachelor, allmenne fag]	325B, 301, 302, 303, 304, 306, 307, 316, 317, 399B
326B Bachelor, humanistiske og estetiske fag	326B, 301, 302, 303, 304, 306, 307, 316, 317, 399B
327B Bachelor, lærerutdanninger og utdanninger i pedagogikk, ikke allmennlærer	327B, 301, 302, 303, 304, 306, 307, 316, 317, 399B
328B Bachelor, samfunnsfag og juridiske fag	328B, 301, 302, 303, 304, 306, 307, 316, 317, 399B
329B Bachelor, økonomiske og administrative fag	329B, 301, 302, 303, 304, 306, 307, 316, 317, 399B
330B Bachelor, naturvitenskapelige fag, håndverksfag og tekniske fag	330B, 301, 302, 303, 304, 306, 307, 316, 317, 399B
331 Høyere nivå's utdanning	*

332 Cand.philol.-utdanning	332, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 316, 317, 318, 319, 320, 321, 322, 323
333 Cand. polit.-utdanning	333, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 316, 317, 318, 319, 320, 321, 322, 323
334 Cand.scient./Cand.real.-utdanning	334, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 316, 317, 318, 319, 320, 321, 322, 323
335 Cand.jur.-utdanning	335, 301, 302
335M Master, rettsvitenskap	335M, 335, 301, 302
336 Cand.med.-utdanning	336, 301, 350
337 Cand.agric.-utdanning	337, 301, 302,
338 Cand.musicae.-utdanning	338, 301, 302, 304, 306, 307, 310, 314, 316, 317, 322, 323, 332, 365, , 318, 319, 320, 321, 322, 323
339 Cand.theol.-utdanning	339, 301, 302
340 Cand.san.-utdanning	340, 301, 302, 303, 304, 305, 306, 307, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323
341 Cand.oecon.-utdanning	341, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 316, 317, 318, 319, 320, 321, 322, 323, 324
342 Cand.psychol.-utdanning	342, 301, 302
343 Cand.sociol.-utdanning	*
344 Cand.sosion.-utdanning	*
345 Cand.act.-utdanning	*
346 Cand.techn.-utdanning	*
347 Cand.pharm.-utdanning	347
347M Master, farmasi-utdanning	347M, 347,302
348 Cand.med.vet.-utdanning	348,
349 Cand.merc.-utdanning	349, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 316, 317, 318, 319, 320, 321, 322, 323, 324
350 Cand.odont.-utdanning	350, 301, 336
350M Master, odontologi-utdanning	301, 336, 350M, 350
351 Cand.ed-/cand.paed./cand.paed.spec.-utdanning	351, 301, 302, 303, 304, 305, 306, 307, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323
352 Hovedfagskandidat	352, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 316, 317, 318, 319, 320, 321, 322, 323
353 Magisterutdanning	353, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 316, 317, 318, 319, 320, 321, 322, 323
354 Siviløkonomutdanning, CEMS-master	*
355 Sivilingeniørutdanning	355, 301, ,
355M Master, teknologifag	301, 355, 355M, 399B, 330B
356 Master of Science	*
357 Master of Philosophy	*
358 Master of Arts	*
359 Master of International Business	*
360 Master of Business Administration	*
361 Master of Management	*
362 Master of Technology Management	*
363 Mastergrader ved statlige høyskoler	*
364 Master of Laws	301, 302, 335, 335M, 364, 364M
365 Annen utdanning på nivå 7, ikke videreutdanning/påbygging	365, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 316, 317, 318, 319, 320, 321, 322, 323
366 Videreutdanninger/påbygging, inntil 4 år	*

367 Spesialistutdanning for leger	*
369M Master, ettårig lavere grad	*
370 Phd.	*
371M Master, allmenne fag	301, 302, 303, 304, 306, 307, 316, 317, 318, 325B, 399B, 371M
372M Master, humanistiske og estetiske fag	301, 302, 303, 304, 306, 307, 316, 317, 318, 326B, 399B, 372M
373M Master, lærerutdanninger og utdanninger i pedagogikk	301, 302, 303, 304, 306, 307, 316, 317, 318, 327B, 399B, 373M
374M Master, samfunnsfag og juridiske fag	301, 302, 303, 304, 306, 307, 316, 317, 318, 328B, 399B, 374M
375M Master, økonomiske og administrative fag	301, 302, 303, 304, 306, 307, 316, 317, 318, 329B, 399B, 375M
376M Master, naturvitenskapelige fag, håndverksfag og tekniske fag	301, 302, 303, 304, 306, 307, 316, 317, 318, 330B, 399B, 376M
377M Master, helse-, sosial- og idrettsfag	301, 302, 303, 304, 306, 307, 316, 317, 318, 396B, 399B, 377M
378M Master, primærnæringsfag	301, 302, 303, 304, 306, 307, 316, 317, 318, 397B, 399B, 378M
379M Master, samferdsels- og sikkerhetsfag og andre sevicefag	301, 302, 303, 304, 306, 307, 316, 317, 318, 398B, 399B, 379M
380M Master, uoppgitt fagfelt	301, 302, 303, 304, 306, 307, 316, 317, 318, 325B, 326B, 327B, 328B, 329B, 330B, 396B, 397B, 398B, 399B, 371M, 372M, 373M, 374M, 375M; 376M, 377M, 378M, 379M, 380M
381 Dr.philos	*
382 Dr.polit	*
383 Dr.scient	*
384 Dr.juris./dr.legis	*
385 Dr.med	*
386 Dr.theol	*
387 Dr.techn	*
388 Dr.odont	*
389 Dr.med.vet	*
390 Dr.agric	*
391 Dr.oecon	*
392 Dr.ing	*
393 Dr.art	*
394 Dr.psychol	*
395 Andre forskerutdanninger (lisensiat)er	*
396B Bachelor, helse-, sosial- og idrettsfag	396B, 301, 302, 303, 304, 306, 307, 316, 317
397B Bachelor, primærnæringsfag	397B, 301, 302, 303, 304, 306, 307, 316, 317
398B Bachelor, samferdsels- og sikkerhetsfag og andre sevicefag	398B, 301, 302, 303, 304, 306, 307, 316, 317
399B Bachelor, uoppgitt fagfelt	399B, 301, 302, 303, 304, 306, 307, 316, 317

* beregnes ikke normert tid for

5.3. Demografitabellen, daterte variable, årlig datering

Demografifilen er en forløpsfil som gir opplysninger om bostedskommune, statsborgerskap og høyeste fullførte utdanning.

5.3.1. Høyeste fullførte utdanning(BHU).

Før omleggingen av registreringen av fullført utdanning i 1999 ble all utdanning på avsluttafilene brukt til å oppdatere BHU. Før omleggingen omfattet avsluttafilene eksamener på minimum 10 vt. i semesteret for universitetene og minimum 5 vekttall i semesteret for høyskolene. Det vil si at tok man for eksempel eksamen på mat-nat fakultetene ble man ikke oppdatert i BHU før man hadde oppnådd

cand.mag fordi de fleste eksamenene der er på mindre enn 10 vt. i semesteret. Altså en betydelig underrapportering på dette fagfeltet. De som derimot har tatt kortere kurs som f.eks. ex.phil blir oppdatert i BHU (per 1.10.98 er det 40 000 personer som står registrert med ex.phil som høyeste fullførte utdanning, og som dermed telles med i andelen personer med høyere utdanning).

Etter omleggingen omfatter avsluttafilene alle avlagte eksamener ved universiteter og høyskoler. Det skilles imidlertid ved hjelp av kjennemerket *recordtype*, mellom

- personer som har fullført en hel utdanning (minimum 2 år) som gir et vitnemål eller en kvalifikasjon, disse har **recordtype = 4**
- de som har tatt kortere kurs eller deksamener som er deler av en utdanning (grunnfag, mellomfag, semesteremner etc.). Disse har **recordtype = 3**.

Dette fører til at rutinene for oppdatering av BHU må endres. F.o.m. 1999-årgangen skal følgende utdanninger oppdateres i BHU:

- ❖ Alle som har fullført
 - ett år utdanning på videregående skole, eller
 - 20 vekttall(60 studiepoeng) grunnutdanning i høgre utdanning, eller
 - 10 vekttall(30 studiepoeng) videreutdanning

Dette skal gjøres på følgende måte for høgre utdanning:

- ❖ Først må 1999-årgangen kodes om til NUS2000 (dette trenger vi bare å gjøre i år, dvs. 1999-årgangen).
- ❖ Så skal det lages en fil der de recordene som skal brukes til å oppdatere BHU skal trekkes ut. Følgende records skal med på denne filen:
 - Alle som har **recordtype = 4** (dvs. har fullført en grad/kvalifikasjon på minst to år), og fullført = 8 og nivå= 6, 7 eller 8.
 - For alle recorder med **recordtype = 3** (dvs. alle typer kortere kurs/deksamener) (inkludert de som ble omkodet i forrige pkt) må det aggregeres vekttall per person for å kunne oppdatere BHU med de som har fullført tilstrekkelig antall vekttall. Det skal bare aggregeres vekttall for de utdanningene som i utgangspunktet ikke er integrerte deler av en sammenhengende utdanning (det er f.eks. ikke noe vits i å oppdatere BHU med de som har tatt to-tre år av legestudiet eller et år av en fireårig ingeniøruddanning). Vi må altså bruke UH-gruppekoden for å plukke ut utdanninger der vi skal aggregere vekttall.

- For "grunnutdanninger" (nivå 6) skal det aggregeres vekttall for records med følgende UHgruppe-kode:

01 Forberedende prøver
02 Lavere nivå's utdanninger
03 Andre ettårige studier, grunnutdanninger

For recorder med **identisk fnr** og verdiene **01-03** i UHgruppe-kode, samt verdien **8** for bestått i *utfomkod* skal antall vekttall i feltet *vektteks* aggregeres opp. De som har til sammen 20 vekttall(60 studiepoeng) legges ut på filen som skal brukes til å oppdatere BHU (jf. recorder med recordtype =4 og kodetype = 2), med en av nuskodene. Denne kan velges tilfeldig, eller man tar den sist avlagte eksamen. Men dersom denne koden har UHgruppe-kode = **01**, eller fagfelt = **0** (2. siffer i nuskoden) velges en av de andre nuskodene.

- For videreutdanninger (nivå 6 eller 7) skal det aggregeres vekttall for records med følgende UHgruppe-kode:

18 Videreutdanninger ledelse/organisasjon/administrasjon/økonomi
19 Videreutdanning for ingeniører
20 Videreutdanning for sykepleiere
21 Videreutdanning for annet helsepersonell enn sykepleier
22 Andre videreutdanninger, til og med to år
23 Praktisk-pedagogisk utdanning (PPU)
66 Videreutdanninger/påbygging, inntil to år
67 Spesialistutdanning for leger

For recorder med **identisk fnr** og verdiene **18-23, 66 og 67** i UHgruppe-kode, samt verdien **8** for bestått i *utfomkod* skal antall vekttall i feltet *vektteks* aggregeres opp **MEN** innenfor hver enkelt av UHgruppekodene. De som har til sammen minst 10 vekttall(30 studiepoeng) innen hver av disse kodene legges på filen som skal oppdatere i BHU, med en av NUS-kodene (dersom det skulle være flere).

- ❖ Denne filen som lages skal altså brukes til å oppdatere BHU etter de tidligere reglene.
- ❖ Det skal etter hvert også gå an å aggregere vekttall fra tidligere år. F.eks om man har tatt noen vekttall i fjor, og med de 10 vekttallene man har tatt i år faktisk har 20 vekttall(60 studiepoeng), bør man oppdateres i BHU. Når vi får utdanningsbasen kan dette gjøres ved å gå inn i denne og se hvor mange vekttall vedkommende har fra de foregående år (med samme UHgruppe-kode). Dette får vi altså ikke gjort første året.

NB!! Det er viktig å være klar over at denne endringen i måten å oppdatere BHU på vil medføre noen endringer i resultatene:

- **Vi vil få med en del eksamener fra mat-nat, spesielt på sikt, ettersom vi aggregerer vekttall over flere år.**
- **Vi vil registrere noen færre eksamener fra de andre fakultetene/høgskolene ettersom grensen for å bli oppdatert i BHU er hevet til 20 vekttall(60 studiepoeng).**

Her er hele programmet for rangering av BHU:

```
SELECT max(
  DECODE(SUBSTR(nus2000,1,1),
    '9','0',
    '0','1',
    '1','2',
    '2','3',
    '3','4',
    '4','5',
    '5','6',
    '6','7',
    '7','8',
    '8','9') ||
  LPAD(TO_CHAR(NVL(kltrinn2000,0)),2,'0') ||
  DECODE(SUBSTR(nus2000,2,1),'0','0','1') ||
  TO_CHAR(regdato) ||
  DECODE(NVL(uhgruppe,'99'),'17','99','23','00','99') )
```

Forklaring:

BHU rangeres etter 1. siffer i NUS2000, der 9 er lavest og 8 er høyest, deretter etter kltrinn2000 der høyeste klasstrinn rangeres høyest, deretter etter 2. siffer i NUS2000 der verdi ulik null rangeres høyere enn verdi lik 0, deretter etter regdato der siste regdato rangeres høyest, deretter etter uhgruppe der verdi ulik 23 rangeres høyere enn verdi lik 23.

Kommentar

Vurdering som ikke resulterte i noen endring.

Det er ikke tatt hensyn til at vi har personer som står med en spesifisert NUS-kode og en detaljert NUS-kode for samme utdanning. Slike uspesifiserte koder har 99 i 5. og 6. siffer i koden. Dette gjelder spesielt koder fra Lånekassa. Når disse i tillegg er hentet fra Helsepersonellregisteret har de en mer spesifisert kode, et eksempel er legeutdanning. Men når vi nå bytter ut koden fra Lånekassa med spesifisert kode fra Helsepersonellregisteret(HPR) før vi legger inn avsluttafilen er det ingen grunn til å ta med dette i valget av BHU-kode. Dersom vi senere velger å legge inn hele HPR i utdanningsdataene bør dette vurderes på nytt.

5.3.2. BU. Ny gruppering av befolkningens utdanningsnivå

For å tilpasse seg de internasjonale standardene og nasjonale interesser på grunn av endringer i utdanningssystemet er definisjonene av befolkningens utdanningsnivå endret. I korte trekk går det ut

på at utdanningnivået er grunnskole inntil hele videregående skole er fullført. Når en person har fullført studiepoeng tilsvarende 2 års studier oppdateres nivået til høyere utdanning. Mer detaljert beskrivelse av endringene finnes i [SSB-magasinet](http://www.ssb.no/vis/magasinet/slik_lever_vi/art-2006-09-14-01.html). (http://www.ssb.no/vis/magasinet/slik_lever_vi/art-2006-09-14-01.html)

Variable i F_UTD_DEMOGRAFI

som knytter seg til gammel BHU:

BHU_NUS2000

BHU_NUS2000DATO

BHU_KLTRINN2000

BHU_KLTRINN2000DATO

BHU_REGDATO

BHU_LOEPENR_KURS

Gammel BHU vil ikke videreføres utover 2005.

Variable som er knyttet til ny gruppering:

BU

BU_DATO

BU_KLTRINN

BU_KLTRINNDATO

BU_REGDATO

BU_LOEPENR_KURS

Ny BU kjøres for alle år fra og med 1976. Tidligere år er BHU og BU identiske.

Ny BU innført i 2006

Avvik mellom gammel BHU og ny BU vil først oppstå for skoleåret 1975/76, dvs BHU 1976.

Spesifikasjonene er laget slik at trinn 2 er avhengig av resultatet av trinn 1, osv.

Trinn 1:

Utdanninger som skal føre til oppdatering:

Alle kurs med rectype ulik 3(4 eller blank) og utfall=8

med unntak av fullført utdanning som ikke skal føre til oppdatering av BU er:

kurs på nivå 3 og klassetrinn 10-11 og avgdato fra og med 197510

kurs på nivå 3 og avgdato fra og med 199510

Trinn 2:

Summerte studiepoeng

UHGRUPPENE 18-23, 66-67:

Tilleggsutdanninger behandles spesielt.

- For videreutdanninger (nivå 6 eller 7) skal det aggregeres studiepoeng for records med følgende UHgruppe-kode:

18 Videreutdanninger ledelse/organisasjon/administrasjon/økonomi
19 Videreutdanning for ingeniører
20 Videreutdanning for sykepleiere
21 Videreutdanning for annet helsepersonell enn sykepleier
22 Andre videreutdanninger, til og med to år
23 Praktisk-pedagogisk utdanning (PPU)
66 Videreutdanninger/påbygging, inntil to år
67 Spesialistutdanning for leger

For recorder med **identisk fnr** og verdiene **18-23, 66 og 67** i UHgruppe-kode, *studpoeng_eks* aggregeres opp, men innenfor hver enkelt av UHgruppekodene. De som har til sammen minst 60 studiepoeng innen hver av disse kodene legges på filen som skal oppdatere i BU, med data fra nyeste record.

Trinn 3:

Ulik UHGRUPPE 18-23, 66-67):

Når summerte studiepoeng er minst 120 skal BHU oppdateres **dersom bu-nivå fra før er ulik 6-8(det må altså oppdateres med grader og tilleggsutdanninger i inneværende år, før det testes på summerte studiepoeng).**

Velg den nyeste recorden som inngår i summeringen som bu-record.

Hvis den valgte recorden har kodetype=1 så skal BU=NUS2000

Hvis den valgte recorden har kodetype=2 så skal BU_NUS = 6||substr(nus2000,2,3)||99.

Forklaring: hvis bu oppdateres på grunnlag av summerte studiepoeng innen en grad/profesjonsutdanning så skal det kodes som uspesifisert innen riktig utdanningsgruppe på nivå 6(fullført 120 studiepoeng av 763101 Cand.med får bu-kode=663199 Medisin, uspesifisert, lavere nivå. Fullført 120 studiepoeng av 661120 Bachelor, sykepleie, treårig får bu-kode=661199 Sykepleiefag, uspesifiserte, lavere nivå).

Trinn 4

Lager en grunnskolerecord i bu når en videregående record ikke fører til oppdatering og personen ikke har en grunnskolerecord i basen.

Dersom en person har

kurs på nivå 3 i kurstabellen og klassetrinn 10-11 og avgdato fra og med 197510 eller

kurs på nivå 3 og avgdato fra og med 199510

og samtidig har BU=999999 eller blank, så skal

BU settes til "201199 Grunnskoleutdanning på ungdomsskoletrinnet, uspesifisert, ungdomsskoleutdanning" og

BU_DATO =FODT_AAR+16

BU_KLTRINN = 00

BU_KLTRINNDATO= FODT_AAR+16

BU_REGDATO= kursets dato fra recorden på nivå 3

BU_LOEPENR_KURS= fra recorden på nivå 3

Det lages også en merkevariabel som angir at det er "laget" en grunnskolerecord.

Kommentarer:

Forberedende prøver

Det ble vurdert om de som bare har fullført forberedende prøverskulle plasseres på videregående-nivå Dette omfatter per 1.10.2005 vel 22000 personer , de fleste er fra før 1999.Det ble bestemt at det i ikke endrese, dvs at de med forberedende prøver som høyeste fullførte blir stående på dette nivået..

Valg av opplysninger ved summerte studiepoengsrecorder

Det velges opplysninger bare fra en 3-er-record ved summerte studiepoeng. Det gjøres for å kunne ha en knytning mellom BU og det kurset i kurstabellen som danner grunnlaget for BU. Det velges siste record av de summerte og det velges dato og NUS-kode fra denne recorden. Siden oppdatering av BU på grunnlag av summerte studiepoeng bare skjer en gang, vil NUS-koden og datoen bli fra den siste eksamenen som er avlagt det første året summen av studiepoeng når 120. Unntaket er summerte studiepoeng fra tilleggsutdanninger, de oppdateres hvert år.

6. Tabellene i NUDB med variabeloversikter

Oversikt over alle tabellene med variabler, definisjoner og kodelister finnes også på <http://www.ssb.no/mikrodata/>

6.1. Demografitabellen. Demografiske opplysninger samt høyeste fullførte utdanning(BHU og BU, gammel og ny definisjon). Forløp med årlig datering per 1. oktober.

- [snr](#)
- [Bostedskommune](#)
- [Statsborgerskap](#)
- [NUS-kode for høyeste fullførte utdanning, gammel definisjon](#)
- [Endringsdato for høyeste fullførte utdanning, gammel definisjon](#)
- [Klassetrinn for høyeste fullførte utdanning\(BHU_NUS2000\) , gammel definisjon](#)
- [Eksakt dato for høyeste fullførte utdanning, gammel definisjon](#)
- [NUS-kode for høyeste utdanning, ny definisjon](#)
- [Endringsdato for høyeste utdanning, ny definisjon](#)
- [Klassetrinn for høyeste utdanning, ny definisjon](#)
- [Eksakt dato for høyeste utdanning, ny definisjon](#)

6.2. Kurstabellen. Forløp med månedlig datering

- [Snr](#)
- [Hovedgruppevariabel](#)
- [Gruppering av skoleslag](#)
- [Utdanningens art NUS2000](#)
- [Utdanningens art NUS89](#)
- [Klassetrinn etter NUS2000-standarden](#)
- [Heltid/deltid](#)
- [Voksenopplæring](#)
- [Kodetype](#)
- [Utfall av utdanning, omkodet](#)
- [Utfall, uomkodet](#)
- [Studieland/fullføringsland](#)

- Poeng ved fullført grunnskole
- Studieretningskode(SSB-kode)
- Kurstrinn i videregående opplæring
- Gruppering av kurstrinn i videregående opplæring
- Kompetanse i videregående opplæring
- Kombinert yrkesfag og allmennfag(TAF)
- Rett til vid.opplæring
- Støtteundervisning i morsmål
- Spesialundervisning etter enkeltvedtak
- Særskilt norskopplæring for språklige minoriteter
- Årstimer for elever under Opplæringsloven
- Delkursprosent for elever og lærlinger
- Dato for realkompetansevurdering i videregående
- Fagopplæringstype
- Elevstatus videregående utdanning
- Fagskoleutdanning
- Gruppering av høyere utdanning
- Etter- og videreutdanning og fjernundervisning
- Opptakskompetanse for studenter
- Med eller uten studierett
- Studiepoeng for grad
- Recordtype på avslutta
- Gjentak
- Studiepoeng for eksamen
- Skolekommune
- Næringskode 72
- Næringskode 78
- Næringskode 94
- Skolens eierforhold
- Høgskolesentra-kode
- Organsisasjonsnummer

- Bedriftens organisasjonsnummer
- Identifikasjonsnr
- Skoletype

6.3. Viktigste kurs. Forløp med månedlig datering

- Studiekode. Første siffer angir hovedgruppen, og resterende siffer angir hvilket studium. Dateres vha. REGDATO
- Tilgangsdato for hovedgruppen.
- Avgangsdato for hovedgruppen.
- Tilgangsdato for studiet.
- Avgangsdato for studiet.
- Utfall, omkodet
- Fagprøvetype
- Antall semester innen studiet.
- Fullført på normert tid innen studiet.
- Antall semester over normert tid innen studiet
- Utdanningens art NUS2000 for fag i semesteret (ikke studiemål)
- SSBs studieretningskode
- Elevens kompetanse etter VKII/ VKIII
- Antall semester registrert i utdanning
- Antall semester i videregående utdanning
- Antall semester i videregående utdanning under lov om videregående opplæring/opplæringsloven
- Antall semester i høyere utdanning
- Antall semester i videregående utdanning i utlandet
- Antall semester i høyere utdanning i utlandet
- Utdannelse i utlandet

6.4. Persontabellen, faste variable på personnivå. (ikke forløp)

- Foreldrenes utd.nivå da personen var 16 år
- Fars høyeste utdanning da personen var 16 år
- Mors høyeste utdanning da personen var 16 år

- Kjønn
- Fødselsår
- Fødselsdato
- Oppholdsdato, første opphold i Norge
- Fødeland
- Innvandringskategori
- Landbakgrunn
- Bostedskommune ved fylte 16 år
- Årstall første gang fullført utdanning på grunnskolenivå
- Årstall første gang fullført tre-årig videregående-skole-nivå
- Årstall første gang fullført tre-årig utdanning under lov om videregående opplæring/opplæringsloven
- Årstall første gang fullført tre-årig utdanning på videregående-skole-nivå, allmennfag
- Årstall første gang fullført tre-årig utdanning under lov om videregående opplæring, allmennfag
- Årstall første gang fullført tre-årig utdanning på videregående-skole-nivå, yrkesfag
- Årstall første gang fullført tre-årig utdanning under lov om videregående opplæring, yrkesfag
- Årstall første gang fullført høgskolekandidat eller tilsvarende
- Årstall første gang fullført cand.mag. eller tilsvarende
- Årstall første gang fullført hovedfagsnivå
- Årstall første gang fullført doktorgrad
- Antall semester registrert i videregående utd. før første fullførte 3-årige videregående utdanning.
- Antall semester registrert i utd. under lov om vid. oppl. før første fullførte 3-årige under loven.
- Antall semester registrert i videregående utdanning før første fullføring, allmennfag.
- Antall semester registrert i vid. utd. under loven før første fullført 3-årig under loven, allmennfag.
- Antall semester registrert i vid. utd. før første fullført 3-årig vid. utd., yrkesfag.
- Antall semester reg. under lov om vid. oppl. før første fullført 3-årig utd.under loven, yrkesfag.
- Antall semester registrert i høyere utdanning før første gang fullført høgskole eller tilsvarende.

- Antall semester registrert i høyere utdanning før første gang fullført cand.mag. eller tilsvarende.
- Antall semester registrert i høyere utdanning før første gang fullført hovedfagsnivå
- Antall semester registrert i høyere utdanning før første gang fullført doktorgrad.
- Antall semester registrert i utdanning utover ordinær grunnskolet før første fullførte 3-årig vid.
- Antall semester reg. i utd. utover ordinær grunnskole før første fullførte 3-årig utd. under loven.
- Antall semester registrert i utdanning utover ordinær grunnskole før første fullføring, allmennfag.
- Antall semester reg. i utd. utover ordinær grunnskole før første fullføring under loven, allmennfag.
- Antall semester reg. i utd. utover ordinær grunnskole før første fullført 3-årig, yrkesfag.
- Antall semester reg. i utd. utover ordinær grunnskole før første fullf. 3-årig under loven, yrkesfag.
- Antall semester utover ordinær grunnskole før første fullført høyskole eller tilsvarende.
- Antall semester registrert i utdanning utover ordinær grunnskole før første gang fullført cand.mag.
- Antall semester registrert i utd. utover ordinær grunnskole før første gang fullført hovedfagsnivå.
- Antall semester registrert i utdanning utover ordinær grunnskole før første gang fullført doktorgrad.
- Årstall første gang registrert i videregående utdanning
- Årstall første gang registrert i videregående utdanning under lov om vid.opplæring/opplæringsloven.
- Årstall første gang registrert i høyere utdanning (ny student)
- År første gang registrert i hovedfag
- Alder ved utgangen av det kalenderåret personen første gang er reg. i høyere utdanning(ny student).

6.5. Karakterer for avsluttet grunnskole

- Statistikknummer
- Organisasjonsnummer
- Skolekommune
- Fagkode for fag i grunnskole

- Standpunktkarakter
- Skriftlig eksamenskarakter
- Muntlig eksamenskarakter
- Avslutningsdato for grunnskolen

6.6. Vitnemål fra videregående utdanning. Data fra Nasjonal vitnemålsdatabase(NVB)

- Snr
- Vitnemålsnummer(skolenr-5, år-4, akkumulator-4)
- Primærvitnemål
- Kurskode for 1. år (UFD-kode)
- Studieretning 1. år (LINDA-kode)
- NUS-kode for 1. år
- Klassestrinn 1. år. SSB
- Studieretning 1. år. SSB
- Kurstrinn 1. år. Detaljert kurstrinn
- Kompetanse i videregående
- Kurstrinn 1. år, gruppert
- Kurskode-2-linje (VK1) UFD-kodeverk
- Studieretning 2. år, detaljert
- NUS-kode 2. år. SSB
- Klassestrinn, 2. året. SSB
- Studieretning, detaljert
- Kurstrinn, 2. året, detaljert
- Kompetanse i videregående, 2. året. SSB
- Kurstrinn 2. året, gruppert. SSB
- Kurskode-3-linje (VK11) UFD kodeverk
- Studieretning, 3. år (LINDA-kodeverk)
- NUS-kode 3. år
- Klassestrinn, 3. år
- Studieretning, 3. år.

- Kurstrinn, 3. år
- Kompetanse i videregående, 3. år
- Kurstrinn, gruppert, 3. år
- Avgangår
- Vitnemålstatus (vitm ikke fyller kravene til bestått vtm)
- Angir om vitnemålet gir generell studiekompetanse
- Orden og atferd
- Fravær-dager år 1 (Antall dager 1. året)
- Fravær-timer år 1 (Antall timer 1. året)
- Fravær-dager år 2 (Antall dager 2. året)
- Fravær-timer år 2 (Antall timer 2. året)
- Fravær-dager år 3 (Antall dager 3. året)
- Fravær-timer år 3 (Antall timer 3. året)
- Utstedelssedato (når vtm ble skrevet ut)
- Utstederskolens organisasjonsnummer
- Skolekommune
- Skolens eierforhold

6.7. Fag fra videregående utdanning. Data fra Nasjonal vitnemålsdatabase(NVB)

- Snr
- Vitnemålsnummer(skolenr-5, år-4, akkumulator-4)
- Fagkode
- Fagets omfang - uketimer
- Standpunktkarakter
- Eksamenskarakter
- Prøveform
- Karaktersnitt for faget
- Eksamensmåned
- Eksamensår
- Fordypningsfag

- Fagtype
- Privatist eller elev
- Organisasjonsnr

6.8. Andre filer ifm NUDB

Fnr-snr-katalog. Brukes for å kunne velge fnr som identifikasjon ved uttak fra basen.
fødselsnummer(Fnr)
statistiknummer(Snr)

Katalog som angir hvilke utdanninger, enten UHGRUPPE-koder eller NUS-koder som det skal telles opp "semester innen studiet" for. Lagret på \$UTD/nudb/arkiv/kat_normert_telles/g2004g2005

Katalog som angir hvilke UHGRUPPER som skal inngå i tellingen av "semester innen studiet". Lagret på \$UTD/nudb/arkiv/kat_normert_tid/g2004g2005

Katalog over fagkoder i NUDB

Katalog over fagkoder i vitnemålsdataene. Lagret på \$UTD/nudb/arkiv/kat_fagkoder/g2001g2006.dat

7. Statistikk- og IT-faglige vurderinger

I arbeidet med å tilrettelegge grunnlagsfilene for forløp i NUDB har alle variabler på de ulike grunnlagsfilene vært nøye gjennomgått for å vurdere deres innhold, kvalitet og nytteverdi. Enkelte variabler har blitt forkastet underveis, mens andre er omarbeidet/omkodet. I dette kapitlet gjør vi en statistikk- og IT-faglig faglig vurdering av de valg som er blitt foretatt på de mest sentrale kjennemerkene i NUDB og problemstillinger knyttet til disse.

7.1. Mulige konsekvenser av påføring av vekttall på eldre filer

Den nye nasjonale utdanningsbasen vil innhente informasjon om hvor mange vekttall/studiepoeng studentene har avlagt. NUS89 inneholder kun et tekstfelt med vekttall for de ulike NUS kodene. Opplysningene om vekttall i NUS89 er ikke tenkt brukt som en tallmessig karakter, men kun som en mer utfyllende beskrivelse for NUS kodene.

Det er ønskelig å legge til vekttall på alle eller deler av NUS89 kodene slik at vi kan påføre vekttall bakover i tid. De eldste filene er ikke omkodet, og inneholder kun NUS73 foreløpig. For å ha et bedre beslutningsgrunnlag før vi vedtar en anbefaling om påføring av vekttall på eldre filer, ble det kjørt en test med et utvalg på 24 personer. Utvalget ble gjort på følgende måte: Et tilfeldig utvalg av NUS89 koder med ulike nivå ble valgt ut. Deretter ble de første og siste personene på avsluttet-filen fra 1998 med disse NUS kodene valgt ut og fulgt i årene 1998-1980. Vekttall ble lagt på manuelt for de ulike NUS kodene ved hjelp av NUS89 katalogen. Resultatene ble lagt inn i en Excel fil for å kunne se totalt antall vekttall for personene i utvalget.

Resultat av testutvalget:

- Personer som har tatt samme fag to ganger får vekttall for faget to ganger.
- Store fagutdanninger som f.eks. sivilingeniør og dr.ing ser ut til å være uproblematisk, i og med at de kun får vekttall ved avsluttet grad.
- Studenter med cand.mag og cand.polit får ikke alltid vekttall for både enkeltksamener, cand.mag, hovedfag og cand.polit, og derfor kan personer med lik utdanning få ulikt antall vekttall. Dette kan skyldes at studentene ikke får vekttall for cand.mag graden før de henter ut vitnemålene sine, og noen venter kanskje helt til de er ferdige cand.polit til de henter det ut. Andre kanskje ikke i det hele tatt.
- Ved kombinasjon av utdanninger kan studenter få for mange vekttall.

Andre problem:

- Tilleggsutdanning for lærere på fra 5-20 vekttall er plassert på samme kode.
- Noen av NUS89 kodene mangler vekttall. I hovedsak gjelder dette doktorgrader, og skal nokså enkelt kunne fylles inn. En løsning er å se hvilke verdier disse kodene har fått i NUS2000, men NUS89 og NUS2000 er bygd opp på ulik måte med flere samlekoder som inneholder mange vekttall.
- Noen utdanninger har endret vekttall over tid. For eksempel var sykepleie- og politiutdanning tidligere på videregående nivå. Vi må derfor undersøke hva som er godkjent for vekttall om de skal føres på eldre filer. Dette kan være en arbeidskrevende jobb.
- Noen NUS koder for lektorer er satt til 80 vekttall. Etter en liten test på disse kodene ser man at enkelte personer med disse NUS kodene ender opp på 80 vekttall, som er for lite, mens andre igjen ender opp med hele 280 vekttall. Dette gjelder imidlertid kun 23 personer.
- Et annet mulig problem er knyttet til studenter med sivilingeniørkoder. Får alle sivilingeniører korrekt antall vekttall eller vil det oppstå feil for de som tar deler av utdanningen på ingeniørhøgskole, og resten ved NTNU. En mulig løsning kan være å sette vekttall på avsluttet grad. Da kan man imidlertid miste eventuelle ekstrarfag disse studentene har avlagt.
- Innrapportering av cand.mag er ikke alltid i samsvar med fullføringsdatao, og dette kan vi ikke gjøre forbedringer på i eldre data.

- Dersom studenter tar kombinasjoner med flere utdanninger. For eksempel må en som tar en sykepleieutdanning og bygger på med noen fag og får cand.mag, ikke få vekttall for både sykepleieutdanningen og cand.mag graden.
- Et annet problem kan være knyttet til studenter som først tar en toårig utdanning, for eksempel en toårig økonomisk administrativ utdanning som gir 40 vekttall, og i tillegg bygger på med ett års videreutdanning innefor revisjon (skal være 20 vekttall i tillegg). Noen av disse studentene vil få registrert det tredje året som siste året i en treårig utdanning, og får 60 vekttall for det siste året. Det er usikkert hvor mange studenter dette kan omfatte.

Konklusjoner og mulige løsninger:

- En løsning er å ikke påføre vekttall på eldre filer, da det er vanskelig og tidkrevende å gjøre dette korrekt. Da vil det bare ligge vekttall fra 1999 årgangen, og vi kan ikke summere opp vekttall tilbake i tid.
- En mulig løsning kan være å kode om til vekttall på kun deler av avsluttet-filene, for eksempel fra 1990 og fram til vi har startet innsamlingen av vekttall. Det må da utarbeides regler for påføring der vekttall mangler. En mulighet er å benytte varighet som kriterium, men enkelte utdanninger har mye praksis som ikke skal være vekttallsgivende. Filene må være merket for gjentak av samme eksamen. Dette kan kanskje løses ved å se forløp i basen. Det må skilles mellom vekttall for grad og vekttall for fag. Er det mulig å skille fullstendig mellom dette? Skal man eventuelt selektene graden med høyeste nivå. (For de med f.eks. cand.mag og cand.polit.). Skal man eventuelt legge til en variabel til som er vekttall for grad minus vekttall for fag, for å se hvor mye studentene har studert i tillegg til graden. Skal man kun selektene ut grad for høyeste nivå? En løsning for kun påføre vekttall for utdanninger som er godkjent for vekttall, er å selektene ut de utdanningene som er på nivå 5 i NUS2000, og kun påføre vekttall på utdanninger som er kodet til klassetrinn 6, 7 og 8 i NUS2000. For å unngå at studenter som tar et år påbygging blir tildelt for mange vekttall kan man kanskje legge inn en test som identifiserer store hopp i antall vekttall på kort tid som ikke skyldes tildeling av en fullført grad.
- Vi kan også velge å legge til vekttall på alle de eldre årgangene. Det vil imidlertid være større feil forbundet med de eldste årgangene, da det vil være en rekke utdanninger som ikke er godkjent for vekttall på de eldste filene, og disse fagene kan være vanskelig og tidkrevende å identifisere. Dette vil involvere at man må gå igjennom alle kataloger for avsluttet-filene for årgangene 79/80 til 97/98 og rette på vekttall, og påføre korrekt antall vekttall hvor det mangler. Vi bør også foreta prøvekoblinger for å se nærmere på resultatet, og se etter urimelige utfall. Regelverk for å unngå feil må utarbeides.

Argument for å kode vekttall på eldre filer:

- Ved lavere grads studier gir ikke klassetrinn god nok informasjon. Vekttall kunne derfor supplere klassetrinn og si noe om varighet av studier.
- Vekttall kan også benyttes for å sannsynliggjøre om det er opphold eller ikke dersom det er "hull" i vårsemesteret (hvor vi ikke har vårsemesterfiler).
- Helheten i basen vil bli styrket med å inkludere vekttall også på filene før 1999.

Konklusjon. Det påføres ikke vekttall/studiepoeng på eldre data:

Vekttallene bør gi mer informasjon en klassetrinn gjør i dag. Dette arbeidet kan bli tidkrevende og gevinsten bør være verdt arbeidsmengden. Konklusjonen blir derfor at det ikke påføres vekttall på eldre data i NUDB, fordi det er komplisert og kvaliteten blir tvilsom.

7.2. Foreldrenes høyeste utdanning/sosial bakgrunn

7.2.1. Foreldrenes utdanning settes en gang

Foreldrenes utdanning er ikke en forløpsvariabel og er kun registrert en gang per person i NUDB.

Foreldrenes utdanning hentes fra demografitabellen i NUDB, det betyr at den er identisk med høyeste utdanning for far og mor 1.10 det året personen fyller 16, eventuelt mors/fars utdanning første gang foreldrene er med i NUDB.

Foreldrenes fødselsnummer hentes fra befolkningsfilen/alle_fnr som inneholder alle personer som har vært registrert siden 1965, totalt 6,5 mill. records.
I NUDB er foreldres utdanning oppgitt for minst 85 prosent av årskullet født etter 1950. For de yngste er uoppgittandelen 2 prosent.

7.2.2. Gruppering av foreldrenes utdanningsnivå

I NUDB tas det inn 3 variabler som belyser foreldrenes utdanning

1. NUS2000 for fars høyeste utdanningsnivå, ny gruppering, 6 siffer (kalt NUS2000_FAR i NUDB)
2. NUS2000 for mors høyeste utdanningsnivå, ny gruppering, 6 siffer (kalt NUS2000_MOR i NUDB)
3. Sosial bakgrunn/foreldrenes utdanningsnivå (kalt SOSBAK i NUDB)
 - 1 = Lang høyere. Mor eller far eller begge har utdanning på nivå 7 el. 8
 - 2 = Kort høyere. Mor eller far eller begge har utdanning på nivå 6
 - 3 = Videregående. Mor eller far eller begge har utdanning på nivå 3, 4 eller 5
 - 4 = Grunnskole. Mor eller far eller begge har utdanning på nivå 0, 1 eller 2
 - 9 = Uoppgitt. Begge foreldrene har uoppgitt utdanning.

I motsetning til hva som er publisert tidligere inneholder NUDB i tillegg til sosial bakgrunn/foreldrenes utdanningsnivå også mors og fars utdanningsnivå på 6-sifternivå. Det ble også vurdert å legge inn foreldrenes utdanning gruppert etter UH-gruppe, men dette ble forkastet. Dersom man er interessert i UH-gruppe for foreldrenes utdanning kan denne enkelt kobles på etter uttak av data fra basen.

Begrunnelsen for å utvide med disse kjennemerkene er at det i framtida vil være like interessant å se sammenhengen mellom egen utdanning og foreldrenes utdanning etter fagfelt som etter nivå.

7.2.3. Bostedskommune

Bostedskommunen har ikke en helt klar definisjon på de opprinnelige utdanningsfilene. For igangværende og avslutta-filene er bosted enten per 1. oktober, per 31.12 eller per det tidspunktet eleven startet i en utdanning/et kurs. En må anta at bostedskommunen på BHU-filene er knyttet til bosted per 1.oktober.

Fordi BHU er mest pålitelig, hentes bostedskommune fra BHU i 1970, 1980, 1985 og 1980-2000. For andre år hentes bostedskommune fra filer fra befolkningsstatistikken.

For årene 1974-1979 og 1981-1984 hentes bosted fra bosattfilene på befolkningsstatistikken. Disse filene er bosatte per 31.12, altså et avvik på 3 måneder i forhold til BHU-filene som er per 1.10.

De aktuelle bosattfilene per 31.12 er nå dokumentert, og viser følgende avvik fra tidligere publiserte tall:

Filen for 31.12.1974 mangler 1668 personer og filen per 31.12.1977 mangler 3228 personer, mens de andre filene har ubetydelige avvik.

Flyttehistoriefilene

Et forslag om å benytte flyttehistoriefilene for å bestemme bosted i i utdanningsbasen har blitt vurdert av prosjektgruppa. Det ble vedtatt å ikke benytte flyttehistoriefilene til å bestemme bosted på elever og studenter i NUDB.

Kommune i basen er det kommunenummeret som gjelder på de aktuelle tidspunktene. BHU-filene har gjennomgått kontroll om at de faktisk inneholder kommunenummer som er gyldige og aktuelle i

miniregisterkontrollen. Det gjøres ingen omkodning av kommunenummer ved kommunesammenslåinger. Dette kan eventuelt gjøres ved kobling mot katalog ved uttak fra basen. Fra 1. november 2001 endres reglene for registrering av bosted for studenter. Etter den nye bestemmelsen kan studenter selv velge om de vil være folkeregistrert i hjemkommunen eller på studiestedet. Tidligere var hovedreglen at studenter ikke skulle melde flytting til folkeregisteret.

7.2.4. Bostedskommune ved 16-års alder eller "hjemstedskommune"

I NUDB er det valgt å legge inn bostedskommunen til personen når han/hun var 16 år. Det er flere grunner til det. Det en enkel variabel som aldri endres og som sier hvor personen kommer fra, og regelverket for registreringen av bosted i Personregisteret er endret fra at studenter var registrert i foreldrenes bostedskommune mens de studerer, med unntak av de som har inngått ekteskap, til at studenter kan velge fritt om de vil være registrert i studiekommunen eller på foreldrenes adresse.

Regler for å bestemme bosted ved 16-års-alder/"hjemstedskommune"

Alle personer født fra og med 1958 påføres bostedskommunen i det året de fyller 16 år. I tillegg får de som er registrert med første oppholdsdato i Norge etter fylte 16 år, bostedskommune 2580 Utlandet. Andre, dvs. de som er født før 1958 blir stående med uoppgitt kommune (kode 9999) ved fylte 16 år. Personer født etter 1957 som fortsatt står med uoppgitt hjemstedskommune, er de som ikke er funnet registrert bosatt det året de fyller 16. Det kan bero på feil i bostedsregistreringen eller at de har vært bosatt i utlandet, men likevel har første oppholdsdato i Norge før fylte 16 år.

I NUDB er det totalt 20311 født etter 1957 med uoppgitt bosted ved 16 år. Av disse er 9839 registrert bosatt i 2006.

7.3. Hoved (hovedgruppe)

Hoved er en tredeling av NUS2000-koden.

0 = Ingen utdanning (nivå 0)

1 = Obligatorisk utdanning (nivå 1-2)

2 = Mellomutdanning (nivå 3-5)

3 = Universitets- og høyskoleutdanning (nivå 6-8)

9 = Uoppgitt (nivå 9)

7.4. Statsborgerskap

Statsborgerskap er hentet fra befolkningsfilene for de årganger der det eksisterer årgangsfiler for befolkningen med påført statsborgerskap (årgangene 1986 - 1989). Årgangsfilene ligger lagret på Unix på flatt format under \$BESTAT/folkem/arkiv/bosatt/g1986m01d01 til g1989m01d01.dat For årene 1990-2000 er statsborgerskap hentet fra de opprinnelige BHU-filene.

7.5. Skoleslag

7.5.1. SSLAG

Variabelen **SSLAG** er en gruppering av skolen og ikke av utdanningen. Skolen er gruppert i henhold til hvordan skolen var definert på det aktuelle tidspunktet.

1 = Grunnskoler

2 = Videregående skoler

3 = Universiteter og høyskoler

Variabelen SSLAG er i de nyere utdanningsdataene erstattet med variabelen UTD som har tre posisjoner (se kapittel 7.5.2). Tanken bak variabelen UTD er muligheten til en grovere inndeling og en mer detaljert inndeling av skoleslag.

Ut fra kjennemerket UTD kan vi lage kjennemerket SSLAG for de nyere utdanningsdataene:

Hvis kjennemerket UTD har verdien 100 **Så**

SSLAG settes lik 1

Hvis kjennemerket UTD har verdiene 211, 212, 220, 510, 520 **Så**

SSLAG settes lik 2

Hvis kjennemerket UTD har verdiene 311, 312, 313, 400 **Så**

SSLAG settes lik 3

7.5.2. UTD

Kjennemerket er lagt inn i NUDB fra og med 198910. Fra og med årgang 1999 benyttes den nye variabelen **UTD** i utdanningsstatistikken. Dette er en variabel som kombinerer gruppering av skoleslag og utdanning, men med vekt på utdanning.

Lærlingene er med i utdanningsstatistikken fra 1989, mens AMO kan skilles ut fra 1987. Fagprøver for skoleåret 1987/88 og tidligere kan ikke merkes i basen uten å gå via teksten i NUS-standardene.

Kodelisten er gjengitt nedenfor.

Kode	Beskrivelse	Igangværende 1989-92 og Avsl.utta 1989/90-1991/92	Igangværende 1993-98 og Avslutta 1992/93-1997/98	Igangværende 1999 og Avslutta 1998/99
100	Grunnskoler	NUS2000 = 2 og (klassebetegnelse ulik 000 og inr som ikke starter på 999) og SN78 < 30	NUS2000 = 2 og (klassebetegnelse ulik 000 og inr som ikke starter på 999) og SN94 < 80.200	Hentes direkte fra filene
211	Videregående skoler	Studieretn. ulik ' ' og (klassebetegnelse ulik 000 og inr som ikke starter på 999) og voksenopplæring = 2	Studieretning ulik ' ' og (klassebetegnelse ulik 000 og inr som ikke starter på 999)	Hentes direkte fra filene
212	Lærlinger	Merke = 'F'	Merke=F	Hentes direkte fra filene
220	Annen videregående utdanning	"resten"	"resten"	Hentes direkte fra filene
311	Statlige høgskoler	Ikke i bruk	Næring = 80.302	Hentes direkte fra filene
312	Militære høgskoler	Ikke i bruk	Næring = 80.303	Hentes direkte fra filene
313	Andre høgskoler	Ikke i bruk	Næring = 80.309	Hentes direkte fra filene
399	Høgskoler	NUS2000 > 5 og SN72 ulik 81	Ikke i bruk	Ikke i bruk
400	Universitet og vitenskapelige høgskoler	SN72 = 81	Næring = 80.301	Hentes direkte fra filene
510	Folkehøgskoler	NUS2000 = 301104, 301105, 301106	NUS2000 = 301104, 301105, 301106	Hentes direkte fra filene
520	Arbeidsmarkedskurs	Klassebetegnelse = 000 og inr som starter på 999	Klassebetegnelse = 000 og inr som starter på 999	Hentes direkte fra filene
610	Utenlandsstudenter videregående			Hentes direkte fra filene
620	Utenlandsstudenter, høyere utdanning			Hentes direkte fra filene
710	Fagskoleutdanning			Hentes direkte fra filene (fra skoleåret 2004/2005)

7.6. Skoleidentifikasjon

Skolen identifiseres ved hjelp av INR, ORGNR og/eller HSKODE

INR er identifikasjonsnummeret fra bedrifts- og foretaksregisteret og er påført dataene fram til og med 1999. Fra 1999 erstattes dette av organisasjonsnummeret(ORGNR). Fra og med skoleåret 2005/2006 er ORGNRBED, som er bedriftens organisasjonsnummer, tatt med i NUDB. Dette for å lette kobling til andre data, for eksempel arbeidsmarkedsdata som også benytter bedriftens organisasjonsnummer.

Fra og med 1994 er det også innført en skoleidentifikasjon for universiteter og høyskoler som har ett felles nummer for alle enhetene som inngår i en statlig høyskole.

7.7. "Manglende vårsemester"

En vårsemesterfil er en fil som inneholder opplysninger om studenter som har startet opp med et kurs på vårsemesteret. Pga. manglende (faglig-)behov for denne filen i utdanningsstatistikken har denne filen ikke vært produsert siden 1998. I oppbyggingen av basen er vårsemesterfilene derfor ikke tatt i bruk. I arbeidet med å bygge opp NUDB har det derimot vært et behov for å få satt en startdato for de studenter som har startet på et kurs i vårsemesteret. Problemet med manglende vårsemesterfil gjelder derimot kun universitetsstudenter(og høyskoler som må behandles på samme måte), altså der UNIV = 1 eller 2. På grunnlagsfilene står alle disse studentene alltid med startdato YYYY09. En student som starter på f.eks. et grunnfag i januar vil av den grunn på miniregisteret alltid stå med startdato YYYY09.

Under danning av forløp i NUDB har vi laget regler som forsøker å bøte på dette problemet. Reglene er som følger:

Hvis NUS2000-koden er den samme som ved forrige registrering **Så**

Dette er samme kurs og startdato fra forrige record gjelder ennå.

Hvis NUS2000-koden ikke er lik forrige registrering **Så**

Hvis studenten var registrert med HOVED=3 på forrige års record **Så**

Hvis studenten var registrert med NUS2000-nivå > 5 høsten før **Så**

Vi antar at studenten også studerte i vårsemesteret, og startdato settes =YYYY01.

Konsekvenser:

Dersom en student var i utdanning på høsten (år YYYY), ikke i utdanning neste vår (år YYYY + 1), men i utdanning neste høst (år YYYY + 1), så vil resultatet bli at vi får studenten registrert også i vårsemesteret selv om han da ikke var i utdanning. Dette er få tilfeller i forhold til de som er i sammenhengende studier, og vi anser at resultatet blir mer korrekt ved at alle blir regnet som studenter i vårsemesteret.

Regel ovenfor gjelder der studenten er registrert ved universitet to høstsemestere etter hverandre, og også dersom studenten var registrert på høyskole første høsten og på universitet neste høst.

7.8. Ktrinn (Klassetrinn gruppert. Gjelder utdanninger under Opplæringsloven/Lov om videregående opplæring/Lov om fagskoler)

Kurstrinn, gruppert. Kodeliste: 1 = Grunnkurs, 2 = Videregående kurs I, 3 = Videregående kurs II/bedriftsopplæring, 4 = Annen utdanning under lov, 5 = Fagskoleutdanning(egen kode fra 2005). Det er kun utdanninger som til enhver tid er en utdanning under Lov om videregående opplæring, under Opplæringsloven eller under Lov om fagskoler som har denne koden.

Grupperer kurstrinn

- Variabelen ktrinn gis verdien 1 (grunnkurs) hvis kurstrin er lik A, B, C, K ,R og kltrinn er lik 11
- Variabelen ktrinn gis verdien 2 (VKI) hvis kurstrin er lik H, J, K, R, I og kltrinn er lik 12

- Variabelen ktrinn gis verdien 3 (VKII) hvis kurstrin er lik P, Q, K, R, I og kltrinn er lik 13
- Variabelen ktrinn gis verdien 4 hvis kurstrin er utfylt; ellers får recordene verdien blank
- Variabelen ktrinn gis verdien 5 hvis kurstrin=U og kltrinn=14 eller 15 eller hvis fagskoleutd=10
- Ellers får variabelen verdien blank

Variabelen **ktrinn** må også legges på de eldre dataene som skal inn i basen. Vi får følgende omkodning for hele datamengden som skal inn i basen (**ktrinn \$char 1**);

- Variabelen ktrinn gis verdien 1 (grunnkurs) hvis kurstrin er lik A, B, C, K, R og kltrinn er lik 11
Variabelen ktrinn gis verdien 1 (grunnkurs) hvis kurstrin er lik 0, 1, 2, 5, 7 og kltrinn er lik 11
Variabelen ktrinn gis verdien 1 (grunnkurs) hvis kurstrin er lik 2 og kltrinn er lik 12
- Variabelen ktrinn gis verdien 2 (VKI) hvis kurstrin er lik H, J, K, R, I og kltrinn er lik 12
Variabelen ktrinn gis verdien 2 (VKI) hvis kurstrin er lik 0, 1, 7 og kltrinn er lik 12
- Variabelen ktrinn gis verdien 3 (VKII) hvis kurstrin er lik P, Q, K, R, I og kltrinn er lik 13
Variabelen ktrinn gis verdien 3 (VKI) hvis kurstrin er lik 0, 1, 7 og kltrinn er lik 13
- Variabelen ktrinn gis verdien 4 hvis kurstrin er lik 0, 1, 7 og kltrinn = 14
Variabelen ktrinn gis verdien 4 hvis kurstrin er utfylt.

7.9. Normert tid i utdanning

Hvorvidt en utdanning er fullført på normert tid eller ikke er en interessant variabel i utdanningssammenheng. Med en slik variabel er det mulig å lage tall for andelen som fullfører en bestemt utdanning på normert tid, 1 år mer enn normert tid, 2 år mer enn normert tid osv., samt sammenligne gjennomsnittlig studietid for de ulike studiene.

Det er ikke alle typer utdanning der det er meningsfylt å se på normert tid. Eksempler kan være mange typer kortere kurs, AMO-kurs, grunnfag, mellomfag o.l. Vi har derfor konsentrert oss om å se på 'hele utdanninger', dvs. treårig videregående skole, fullførte grader, profesjonsutdanninger o.l.

For å lage indikatorer for fullføring på normert tid, har vi definert normert tid for de enkelte utdanningene. Det finnes variabler for varighet knyttet til nus2000 koden. For høyere utdanning (nivå 6, 7 og 8) er varigheten oppgitt i antall vekttall. For annen utdanning er varigheten oppgitt i måneder. Problemet med disse er at de ikke er kumulative, dvs., for utdanninger som består av flere deler er det bare oppgitt varigheten for den aktuelle delen (f.eks. hovedfag/cand.polit = 40 vekttall).

For å sette normert tid er det utarbeidet en katalog med normert antall semestre i utdanning sammenlagt for de ulike utdanningene.

For videregående utdanning (nivå 3 og 4) er det tatt utgangspunkt i Kompetansekoden. Kompetansekoden er kun påført fra og med 1997 og kun på utdanninger som omfattes av Reform94. *(De som har endret KOMP underveis vil bli stående med den koden som gjaldt da kurset startet. Problemet er de som startet før 1997, de vil stå med blank i KOMP fordi komp ikke var med på filene da kurset startet. Men disse må ha startet i lære før R94 og da skal de antagelig ikke ha KOMP. Fra NUDB kan KOMP benyttes for avslutta utdanning først fra og med 2000/2001 og for igangværende fra 2000)* For de elever som ikke har kompetansekode beregnes normert tid på følgende måte:

Fra og med avslutta 1988/89 og i gang 1989:

Hvis kompetansekode er påført Så

Normert varighet settes ut fra katalog

Ellers

Hvis kompetansekode = blank Og kurstrinn er utfylt Og klasstrinn = 13 Og UTD ulik 212 **Så**
Normert varighet = 6 semester

Hvis kompetansekode = blank Og kurstrinn er utfylt Og klasstrinn = 14 Og UTD ulik 212 **Så**
Normert varighet = 8 semester. (vkIII)

Hvis kompetansekode = blank Og klasstrinn = 13 Og UTD =212 **Så**
Normert varighet = 8 semester. (fagopplæring)

Hvis kompetansekode = blank Og klasstrinn = 14 Og UTD =212 **Så**
Normert varighet = 10 semester. (fagopplæring)

Ellers skal det ikke settes noen verdi for normert tid

- For høyere utdanning (nivå 6, 7 og 8), er normert tid satt i katalogen ut fra uh-gruppekoden.

Lærlinger kom med i statistikken først fra 1.10.1989.

I lære defineres som UTD = 212 Lærlinger/fagprøver. Kjennebraket UTD er første gang med i NUDB fra iganghværende 1989. (er datert)

- År første gang fullført utdanning på videregående-skole-nivå
 - bestått (B, 8)
 - ikke bestått (I, A)

7.9.1. Telling av antall semester i NUDB

Telling av antall år/semester under utdanning.

Det er flere variabler i Nudb som går på antall år i utdanning, på forskjellige nivåer, i utdanning totalt, utover normert tid osv. Hvordan antall år i utdanning skal defineres må imidlertid spesifiseres nærmere. Følgende alternativer er vurdert:

1. Telle antall ganger en person er registrert som igangværende per 01.10 eller
2. Telle antall måneder en person er under utdanning og dele på et forholdstall (8 eller 10 (hhv. høyskole eller universitet) dersom utdanningen tas under høstsemesteret, og 12 dersom utdanningen tas under vårsemesteret). Resultatet vil bli en brøk som vi kan runde av til nærmeste halve år (bør vi i tillegg ta vare på den eksakte brøken for å kunne summere flere delutdanninger riktig på et senere tidspunkt dersom personen har opphold mellom kursene, eller bør vi...)
3. Telle antall semester som en person har vært registrert med kurs i. Et ½ år settes lik et semester

Velges alternativ 1. sier vi at dersom en person er registrert i utdanning 1.10 et gitt år teller dette som ett år i utdanning. Dette er den enkleste måten å gjøre det på, og også den som er den som baserer seg på de sikreste dataene. Vi vet ganske nøyaktig hvem som var i utdanning 1.10 fra filene over igangværende utdanning. Det er imidlertid ikke sikkert at alle som var igangværende 1.10 i høstsemesteret også er i utdanning i vårsemesteret. Denne måten å telle antall år under utdanning er sånn sett lite nøyaktig.

Alternativ 2 er en mer detaljert fremgangsmåte, der antall måneder i utdanning skal summeres. Denne måten å telle antall år i utdanning er ganske nøyaktig dersom man forutsetter at alle datoene på filene er korrekte. Vi vet imidlertid at dette ikke er tilfelle. Det er mye feil i datoer på utdanningsfilene, og disse blir i tillegg endret etter visse regler ved danning av miniregistre. Denne måten å telle antall år under utdanning vil derfor være litt for detaljert, tatt i betraktning kvaliteten på datoene på filene.

Alternativ 3 er en mellomting mellom de to første alternativene når det gjelder detaljeringsgrad. Her telles alle semesterne en student har vært registrert i gang med en utdanning, uansett varighet. Vårsemester = jan-juli, høstsemester = aug-des. Har man en eller annen registrering i et semester (start/i gang/avslutning) telles semesteret, uansett dato. Denne måten å telle år i utdanning er mer nøyaktig enn alt. 1, men mindre detaljert enn alt. 2. Antall semester blir liggende i basen, men kan regnes om til år i tabeller der det er ønskelig.

Konklusjon:

Alternativ 3 ser ut til å være den beste løsningen. Dette alternativet vil gi et forholdsvis riktig bilde av antall semester i utdanning, det er gode grunner for å si at dersom man har en eller annen registrering av utdanning i et semester har man vært i gang med utdanning dette semesteret.

Telling av antall semester i utdanning i NUDB:

Variablene SEMESTER_I_VUTD og SEMESTER_I_HOY_UTD blir telt opp parallellt, slik at antall semester i utdanning totalt IKKE er summen av semester_i_vutd og semester_i_hoy_utt. For studenter i utlandet er det motsatt, dvs. slik at det telles enten i utlandet eller i Norge.

7.10. Definisjon av semester

Generelt sier vi at vårsemester = januar-juli, høstsemester = august-desember.

Men et kurs som går i sommersemesteret, f.eks. fra april til august vil jo strekke seg over to semester, og kan dermed bli telt som ett år. Det samme med kurs som strekker seg over jul/nyttår. For å unngå dette må slike tilfeller behandles spesielt.

- Dersom et kurs strekker seg over to semestre, men varer fem måneder, eller kortere, skal det telles som ett semester. Dersom kurs et varer mer enn fem måneder skal det telles som to (eller flere). Generelt kan vi si at utdanning som strekker seg over sommeren skal høre med til vårsemesteret. Ellers puttes utdanningen til det semesteret det strekker seg mest inn i.

Datering av variabelen 'antall semester i utdanning'

Oppdatering av basen vil skje en gang i året, på høsten etter at årgangsfilene for utdanning er ferdige. Opplysningene om utdanning er per 1.10, og det vil derfor være naturlig at datoen for 'antall semester i utdanning' og lignende variabler også settes til 1.10.

7.11. Gjennomstrømningsvariable i NUDB

For å legge til rette for enklere uttak av forløp i NUDB er det laget to typer variabler som beskriver gjennomstrømning på ulike måter. Den første typen er kjennemerker som settes en gang og som det ikke er forløp på. Den andre typen er forløpsvariabler som kan endres over tid, og som derfor dateres.

7.11.1. Gjennomstrømningsvariable som settes en gang/ved 1. gangs hendelse. (TAB_UTD_PERSON))

Første gang fullført

1. Årstall eller årstall og måned 1. gang fullført
2. Antall semester i høyere (el. videregående el. under lov om videregående) utdanning før 1. gang fullført
3. Antall semester i utdanning totalt før 1. gang fullført

Disse tre variablene gjentas for følgende grupperinger:

- 3-årig videregående utdanning
- 3-årig vid. utdanning, allmennfag
- 3-årig vid. utdanning, yrkesfag

- 3-årig utd. under lov om videregående opplæring
- 3-årig utd. under lov om videregående opplæring, allmennfaglige studieretninger
- 3-årig utd. under lov om videregående opplæring, yrkesfaglige studieretninger

- høgskolekandidat eller tilsvarende
- cand.mag.-nivå eller tilsvarende

- hovedfagsnivå
 - doktorgradsnivå
- Disse knytter seg alle til 1. gang fullført og skal ligge i TAB_UTD_PERSON.

Andre gjennomstrømningsvariable i TAB_UTD_PERSON

- Årstall 1. gang fullført grunnskole
- Årstall 1. gang registrert i videregående utdanning etter grunnskolen
- Årstall 1. gang registrert i utdanning under lov om videregående opplæring etter grunnskolen
- Årstall ny student/1. gang registrert i høyere utdanning
- Alder ny student/1. gang registrert i høyere utdanning

Ved hjelp av årstall for start på ulike nivåer og årstall for fullført på ulike nivåer kan det enkelt regnes ut antall år før fullført ulike utdanninger (da medregnet opphold).

7.11.2. Gjennomstrømningsvariable som det dannes forløp på. (I F_UTD_PERSON)

- antall semester i utdanning totalt
- antall semester i videregående utdanning totalt
- antall semester i videregående utdanning under lov om vid. opplæring/opplæringsloven
- antall semester i høyere utdanning totalt

I de neste variablene refererer antall semester seg til semester innen vedkommende utdanning:

- antall semester innen vedkommende utdanning (for å kunne lage tabell med gjennomsnittlig studietid)
- fullført på normert tid
 - 1=under normert tid, 2=på normert tid, 3=mer enn normert tid
- antall semester utover normert tid i vedkommende utdanning.

7.12. Valg av identifiserende nøkkel i NUDB

7.12.1. Oversikt over nøkkelalternativene

NUDB omfatter mange ulike grunnlagsregister over et stort tidsrom. Registerne omfatter bl.a. alle utdanningsfiler (igangværende og avslutta utdanning) samt BHU-filer fra 1970 og fram til i dag. I NUDB (og andre statistikkssystemer) er det essensielt at den nøkkelen som benyttes for å identifisere enehetene, f.eks. en person, er **unikt identifiserende**. I grunnlagsregisterne som er benyttet for å bygge opp NUDB er fødselsnummer benyttet som en slik unikt identifiserende nøkkel på personnivå. Som vi skal se har fødselsnummer en del egenskaper som gjør at det ikke er helt trivielt å benyttes fødselsnummer for dette formålet.

7.12.1.1. Fødselsnummer

Av ulike årsaker vil en rekke personer i løpet av en gitt tidsperiode ha byttet fødselsnummer, noen flere ganger, underveis. Punktene nedenfor er en oversikt over hva som kan ligge bak de byttene som skjer:

1. Person har byttet kjønn, eller kjønn ble feil registrert i utgangspunktet
2. Person har endret fødselsdato, eller fødselsdato ble feil registrert i utgangspunktet
3. Dubletter. Samme person har hatt to ulike Fnr og to ulike Snr. F.eks. kan en person som har vært registrert bosatt i Norge i flere perioder, ha fått tildelt flere ulike fødselsnummer med tilhørende ulike Snr.
4. Person har fått tildelt D-nummer først, for deretter å få et gyldig Fnr

Generelt kan en derfor si at fødselsnummer er unikt identifiserende for person **innenfor hvert register**, men ikke mellom ulike register (f.eks. samme register ulik årgang/ ulik register samme årgang/ ulike statistikkområder etc.). En konsekvens av dette er at vi ved kobling av flere register innenfor utdanning (eller utdanning mot andre statistikkområder) ikke vil oppnå en optimal

treffprosent. I tillegg vil det være en fare for at samme person telles opp flere ganger, eller at vi danner flere ulike forløp for en og samme person.

7.12.1.2. Snr (statistiknummer)

Det er i SSB opprettet en katalog som inneholder alle gyldige fødselsnummer som er tildelt fra 1964 og fram til i dag. I katalogen er alle tidligere tildelte fødselsnummer for hver person linket sammen vha. et Snr. På denne måten er Snr i motsetning til fødselsnummer, unikt identifiserende for person **på tvers av registrene** siden hver person i Snr katalogen vil ha sin egen unike id, uavhengig av hvor mange ganger personen har skiftet fødselsnummer.

7.12.2. Valg av løsning

Siden fødselsnummer er en sentral del av nøkkelen på alle nivåer i NUDB, har det vært viktig å finne løsninger som ivaretar og sikrer konsistens over tid i datavarehuset. Det finnes per i dag to (tre) kataloger/databaser som kan benyttes for å sikre at vi oppnår ønsket konsistens i NUDB. Nedenfor er det gitt en kort beskrivelse av hver av disse samt en matrise med oversikt over hvor godt hver enkelt av dem er egnet til å løse de fire problemområdene som er nevnt i kapitell 7.12.1.1 om fødselsnummer.

7.12.2.1. Snr-katalogen

Snr-katalogen oppdateres en gang i måneden og inneholder følgende variable:

- Fnr (fødselsnummer, dvs. alle fødselsnummer)
- Fnr_naa (fødselsnummer som gjelder nå)
- Snr (et entydig statistisk nummer)
- dato_fnr (dato da fødselsnummer ble opprettet)
- dato_fn2 (dato da person sist endret fødselsnummer)

7.12.2.2. BEBAS

Filen som benyttes er en kopi av BEBAS, og den omfatter følgende 3 variable:

- alle Fnr som har vært i bruk
- tidligere Fnr (dublett)
- nytt Fnr

BEBAS inneholder alle tidligere Fnr med link til det neste (påfølgende) Fnr. Dubletter (at en person har fått tildelt to eller flere fødselsnummer (og Snr) i ulike perioder) kan også finnes vha. BEBAS, men per i dag finnes det ingen egentlig LINK-fil for disse dublettene.

For å finne fram til siste Fnr må en koble på nytt inntil ingen Fnr har nytt Fnr. En LINK-fil må inneholde følgende to variable:

- alle Fnr som har vært i bruk
- nyeste Fnr.

7.12.2.3. Fil med link mellom Dnr. og Fnr.

D-nr er et nummer som tildeles personer som har forbindelse med norske myndigheter av økonomisk art, men som ikke tilfredsstiller kravene for å få fødselsnummer. Mer offisiell forklaring finnes i D-nummerforskriften. Datoene er påplussert 40.

Linken mellom Dnr. og Fnr. foreligger som en egen fil, og er verken en del av Snr-katalogen eller BEBAS. Antall med Dnr. i utdanningsdataene er antakelig svært lavt, i alle fall inntil dags dato, fordi vi har endret alle ugyldige Fnr (les: ugyldige Fnr og Dnr) til kunstige Fnr.

7.12.2.4. Linkmuligheter i de forskjellige kataloger/databaser:

Årsaker bak endret fødselsnummer	Antall personer over 16 år	Snr-katalogen	BEBAS	Fil med link mellom Dnr og Fnr
1. Person har byttet kjønn, eller kjønn ble feil registrert i utgangspunktet	ca. 50 i løpet av siste år.	Det finnes en entydig link mellom alle gamle fødselsnummer og det nye fødselsnummeret	BEBAS har link mellom tidligere Fnr og neste Fnr, men ikke mellom alle tidligere Fnr og siste Fnr. For å finne fram til gjeldende Fnr per dags dato må en koble på nytt mot en LINK-file (som må opprettes) inntil ingen Fnr har nytt Fnr. Dette må gjøres for hvert register som legges inn i NUDB, og for hver gang det gjøres uttak for kobling mot andre statistikkområder.	Ikke aktuelt
2. Person har endret fødselsdato, eller fødselsdato ble feil registrert i utgangspunktet	ca. 280 i løpet av siste år.	Det finnes en entydig link mellom alle gamle fødselsnummer og det nye fødselsnummeret	BEBAS har link mellom tidligere Fnr og neste Fnr, men ikke mellom alle tidligere Fnr og siste Fnr. For å finne fram til gjeldende Fnr per dags dato må en koble på nytt mot en LINK-file (som må opprettes) inntil ingen Fnr har nytt Fnr. Dette må gjøres for hvert register som legges inn i NUDB, og for hver gang det gjøres uttak for kobling mot andre statistikkområder.	Ikke aktuelt
3. Dubletter. Samme person har hatt to ulike Fnr og to ulike Snr	Ca. 20-25 i løpet av siste år	Det finnes per i dag ingen link i Snr-katalogen mellom disse fødselsnummerne og Snr. Det er mulig å legge dette inn som en del av Snr.-katalogen, men det er en "større" jobb.	Det må opprettes en LINK-fil i BEBAS hvor dublettene er linket sammen. Denne LINK-filen må ajourføres jevnlig for å være oppdatert.	Ikke aktuelt
4. Person har fått tildelt Dnr først, for deretter å få et gyldig Fnr	Svært få før år 2000 da de fleste er kodet om til ugyldige Fnr på utdanningsfilene.	Det finnes per i dag ingen link i Snr-katalogen mellom disse Dnr og Snr. En slik link bør vurderes dersom en bestemmer seg for å gjøre noe med Dnr på utdanningsfilene/BHU i årene framover.	Det finnes per i dag ingen link i BEBAS mellom disse Dnrene og Fnr. En slik link bør vurderes dersom en bestemmer seg for å gjøre noe med Dnr på utdanningsfilene/BHU i årene framover	Har link mellom alle Dnr som har fått tildelt gyldig Fnr på et eller annet tidspunkt.

7.12.3. Implementering av nøklene i NUDB

Som nevnt tidligere er NUDB er bygd opp av to ulike nivåer; ett som følger personens kurs (KURS-nivå) og ett som følger personens utdanningsnivå (PERSON-nivå). I tillegg er det et forløpsnivå med demografiopplysninger. For alle nivåer vil fødselsnummer være en viktig del av den unike identifiserende nøkkelen for en record. Nedenfor er en gjennomgang av de fire ulike løsningsforslagene med konsekvensanalyser som ble vurdert underveis.

7.12.3.1. Benytte BEBAS/LINK-filen ved uttak

Forkastet løsningsforslag:

Legger fødselsnumrene inn i NUDB slik de er på utdanningsfilene/BHU, og benytter BEBAS/LINK-filen som katalog for å påføre det siste gyldige fødselsnummer ved uttak.

Konsekvens:

Medfører at vi kan risikere å danne ulike parallelle forløp i NUDB for en og samme person og at dette ikke blir oppdaget før ved uttak fra basen hvor det er svært vanskelig/tidkrevende å gjøre noe med det. Det vil heller ikke være mulig å telle opp variable i NUDB uten å ta ut data via BEBAS/LINK-filen først (mao. vil det ikke være mulig å gi korrekte svar ved ad-hoc spøringer mot NUDB). Ved kryptering vil i tillegg hvert eneste uttak medføre ekstra tidsforbruk for å få koblet på riktige fnr.

7.12.3.2. Benytte Snr-katalogen ved uttak

Forkastet løsningsforslag:

Legger fødselsnumrene inn i NUDB slik de er på utdanningsfilene/BHU, og benytter Snr-katalogen for å påføre det siste gyldige fødselsnummer ved uttak

Konsekvens:

Samme som pkt. 7.12.3.1

7.12.3.3. Løpende endring av Fnr i NUDB, og ved uttak.

Forkastet løsningsforslag:

Vi endrer fødselsnummer i NUDB underveis mens vi bygger opp datavarehuset: Med andre ord bytter vi ut en del av den unike nøkkelen for en record etter at den er benyttet for å bygge opp forløpet. I tillegg må vi koble mot BEBAS/LINK-fil ved uttak for å sikre oss de siste gyldige Fnr.

Konsekvens:

Løsningen innebærer at rutinenene for å danne NUDB blir atskillig mer komplisert (hver person i hvert register må matches mot en omkodingstabell før de matches mot personer i NUDB). Ved match mellom omkodingstabellen og ny person fra et register vil vi måtte løpe gjennom alle record for denne personen i NUDB og erstatte det gamle fødselsnummeret med det nye.

Siden BEBAS/LINK-filen endrer seg fortløpende vil vi også måtte koble alle uttak fra NUDB mot BEBAS/LINK-filen for å få rett opp ev. endringer i Fnr som er skjedd etter at forløpet ble dannet. Dette medfører at vi kan risikere å ha dannet ulike parallelle forløp i NUDB for en og samme person og at dette ikke blir oppdaget før ved uttak fra basen hvor det er svært vanskelig/tidkrevende å gjøre noe med det. I praksis betyr dette at det heller ikke ved denne løsningen vil være mulig å telle opp variable i NUDB uten å ta ut data via BEBAS/LINK-filen først (mao. vil det ikke være mulig å gi korrekte svar ved ad-hoc spøringer mot NUDB). Ved kryptering eller ved uttak når vi ønsker å koble mot andre databaser/register, må det kobles mot Snr-katalogen først for å få påført de riktige Snr. Uten Snr vil ikke koblinger mot andre statistikkområder/egne filer gi eksakt match.

I sum vil hvert eneste uttak av data fra NUDB medføre betydelig ekstra tidsforbruk for å få koblet på riktige Fnr. Enda mer tidkrevende og komplisert vil innleggingen av nye data i NUDB være.

7.12.3.4. Benytte Snr som identifiserende id for personer i NUDB

Godkjent løsningsforslag:

Vi kutter ut bruken av fødselsnummer I NUDB. Data for personer som legges inn i NUDB blir istedenfort identifiseres vha. Snr (i kombinasjon med andre variable). Dette betyr at det er Snr og ikke fødselsnummer som benyttes både ved kobling mellom ulike grunnlagsregistre ved oppbygging av NUDB, og ved kobling mellom NUDB og andre statistikkssystemer senere. Det er imidlertid lagt opp til at fødselsnummer kan påføres ved uttak av data fra NUDB der det er ønskelig/påkrevd. Da det kun er de 8 første posisjonene av Snr som er unike (de tre siste angir kjønn og fødselsår), er det disse som er benyttet i NUDB.

Konsekvens: Ved å ta i bruk Snr allerede i databasen har vi sikret konsistente data (entydige forløp, med unntak av dublettene som er beskrevet i kapittel 7.12.1.1, punkt 3) i NUDB med minimal innsats. Det er også blitt betydelig lettere å koble utdanningsdata mot andre statistikkområder samtidig som vi har sikret at alle record som skal matche ved en kobling faktisk gjør det. Da Snr kun er et løpenummer og et mellomsteg for data som skal krypteres, vil data i NUDB være sikret mot ev. misbruk på et helt annen nivå enn ved bruk av Fnr.

For uttak av data for statistikkproduksjon og analyse vil det være behov for å kunne ta ut data fra NUDB med påført Fnr istedenfor Snr. Denne opsjonen er gjort tilgjengelig i uttakssystemet for NUDB. Oppslag direkte mot NUDB for å se på enkelttilfeller (for test av datamaterialet) gjøres enten vha. en SQL-spørring hvor det kobles mot Snr-katalogen (les: Oracle-tabell med kobling mellom fødselsnummer og Snr), eller vha. programmet FoFI_NUDB (visualiseringsprogram for forløpsdata i NUDB).

For fag-siden vil de praktiske konsekvensene av å benytte Snr bli minimale.

Dersom en ønsker å gjøre noe med dublettene og Dnr som beskrevet i kapitlene over, vil en innsats mot Snr-katalogen kunne løse dette med relativt beskjedent tidsforbruk, sammenlignet med tidsforbruket for å få implementert de andre løsningsalternativene.

7.12.4. Ugyldige fødselsnummer og Dnr

Ugyldige fødselsnummer er lagt inn som de er, men det er ikke laget forløp på ugyldige fødselsnummer i NUDB. De er likevel tatt med i NUDB for å kunne reprodusere tidligere publiserte tverrsnitt på BHU. Når det gjelder Dnr er disse ikke lagt inn i NUDB. De er estattet med kunstige fnr, og blir behandlet som disse. Ugyldige fødselsnummer som er lagt inn i NUDB har fått prefixet 'U' ('Uxxxxxxx'; hvor xxxxxxx er et løpenummer).

7.13. Studieretning 13 og 15

Studieretning er første gang publisert i 1976. Det er en gruppering som er knyttet til Lov om videregående opplæring, som ble innført i 1976. For årene 1974-1975 har grunnlagsfilene påført studieretningskode, selv om denne aldri er publisert. Vi har valgt å beholde disse kodene i NUDB, men har satt studieretning 13 og 15 til blank, da det ikke har vært mulig å finne dokumentasjon av disse. Ved kontroll finner vi dessuten at de samme kursene som står med studieretning 13 og 15 i 74-75 er satt til blank i studieretning i 1976.

Dette er løst på følgende måte: Alle med studieretning 13 og 15 gis blank i studieretning og kurstype, og i alle variable som er avledet fra disse.

8. Feil og mangler i dataene. Endringer i prinsipper, definisjoner og avgrensning av massen

8.1. Innledning

Seksjon for utdanningsstatistikk har utviklet en nasjonal utdanningsdatabase (NUDB) som omfatter utdanningsdata på individnivå fra og med 1970 når det gjelder befolkningens utdanningsnivå, fra og med 1974 når det gjelder igangværende utdanninger og fra og med 1974/75 når det gjelder fullførte utdanninger. Det er et meget omfattende datamateriale som inngår i basen. Materialet omfatter data fra grunnskolenivå til forskernivå, da unntatt igangværende utdanninger i grunnskolen. Datamaterialet gir unike muligheter til å publisere ulike former for utdanningsstatistikk. Ikke minst er basen særdeles viktig for å produsere gjennomstrømningsdata. Slike data er i dag sentrale som grunnlag for mye av den pågående utdanningsforskningen og for den offentlige planleggingen og evalueringen av utdanningspolitikken. Datakildene har variert over tid, men dataene er i hovedsak blitt hentet fra det enkelte lærested og fra ulike administrative registre. Datafangsten har selvfølgelig endret seg en del underveis. Ulike registre har etter hvert fått større og større betydning.

Kvaliteten på dataene i NUDB er gjennomgående god. Det er brukt mye arbeid på kvalitetssikring. NUDBs kvalitet avhenger av kvaliteten på de utdanningsdataene som er hentet inn hvert år, og den bearbeiding av dataene som er gjort i SSB fram til ferdigstilling av de ulike årgangsfiler som er blitt benyttet i basen. Med et så stort datamateriale som er hentet inn over så mange år og som har vært gjennom så omfattende bearbeidingsprosesser er det ikke til å unngå at det vil bli enkelte feil og mangler. I denne oversikten vil vi redegjøre for de viktigste av disse som vi kjenner til. Det finnes ingen skriftlig dokumentasjon om kvaliteten på årgangsfilene hvert år. Det er derfor helt umulig å ha en fullstendig oversikt. Det er imidlertid viktig at feil og mangler ved dataene som vi er kjent med og som kan få betydning for resultater av forskning og studier, også er kjent for brukerne av dataene i NUDB.

Like viktig som å redegjøre for feil og mangler er det å gi en oversikt over endringer i prinsipper, definisjoner og omfang av dataene som inngår i datamassen hvert år. Når det er data som er innhentet over et så langt tidsrom vil det nødvendigvis måtte bli en del endringer. Ikke minst vil endringer i utdanningsstrukturer ha stor betydning. For brukere av NUDB er derfor denne oversikten viktig.

8.2. Feil og mangler

8.2.1. Generelt

Det kan være ulike feilkilder i statistiske undersøkelser. Feil kan oppstå ved innsamling av data ved at de enheter vi undersøker ikke alltid er helt identiske med den masse av enheter vi ønsker å undersøke. Årsaken kan være at vi ikke kjenner til alle aktuelle oppgavegivere og derfor ikke får samlet inn alle data eller at vi ikke får data fra alle oppgavegivere vi kjenner til. En oppgavegiver kan også gi feilaktige eller mangelfulle opplysninger. Av og til vil det også oppstå feil i bearbeidingen av data som er kommet til SSB, f.eks. kodefeil.

Det registrerte tallet på studenter hvert år må vies spesiell oppmerksomhet. I studentstatistikken vil det være flest feil og mangler i statistikken der datagrunnlaget bygger på opplysninger gitt av den enkelte student. Grunnlaget for studenttallene har i alle år vært tallet på registrerte studenter ved det enkelte lærested. Tallet på registrerte studenter vil derfor avvike noe fra det reelle studenttallet. Registrerte studenter vil si studenter som har betalt semesteravgift/studieavgift.

Det har i årenes løp både vært en overrapportering og en underrapportering av studentdata. Underrapportering er nok det minste problemet. Dette gjelder i første rekke fjernundervisningsstudenter som ikke er registrert som studenter de semestrene de ikke skal avlegge

noen eksamen. De kan likevel være aktive studenter. I eksamenssemesteret må også disse studentene betale semesteravgift og dermed registrere seg, så i eksamenssemesterene vil alle rapporteres. Dette problemet har vært størst fra og med slutten av 1980-tallet da tallet på fjernundervisningsstudenter økte. Det er vanskelig å anslå hvor omfattende denne underrapporteringen av reelle studenter er.

Et større problem er overrapportering av studenter. Det er en del studenter som melder seg på studier og betaler semesteravgift som aldri avlegger noen eksamen, som stadig utsetter sine eksamener eller som i realiteten har avsluttet sine studier. Dette gjelder spesielt universitetsstudentene. Det kan være ulike årsaker til at studenter registrerer seg på studier og melder seg opp til eksamener uten å være aktive studenter på heltid eller deltid. For noen kan det være at jobb prioriteres foran studier, for andre kan det være sykdom eller at man er indisponert for studier i kortere eller lengre perioder. Dessuten lønner det seg økonomisk for mange å registrere seg som student, betale semesteravgift og få utstedt studiebevis. Dette gir krav på en rekke rabatter ved kjøp av ulike tjenester. Enkelte undersøkelser kan tyde på at det er relativt mange slike "studenter" uten at det er mulig å si eksakt hvor stort antall disse fiktive studentene faktisk utgjør. Ved å samkjøre registeropplysninger om sysselsetting og utdanning viser det seg at f.eks. en god del av de som avslutter en lang høyere utdanning fortsetter å være registrert som student selv om de er sysselsatt på fulltid. Registreringsproblemene har tradisjonelt vært minst for de mest fast strukturerte profesjonsstudiene og størst for de mer løst oppbygde studier.

Det er ikke mulig å anslå hvor stor overrapporteringen av studenter har vært i hele perioden fra 1974 og fram til i dag, dvs. i hele den perioden vi har individdata. Fordelen ved å være registrert som student har variert og vil i alle fall til en viss grad påvirke dette tallet. Generelt kan vi nok fastslå at problemet har vært størst de siste årene. Etter hvert er det blitt svært mange studenter, og veldig mange av disse studentene er ikke heltidsstudenter. Bildet er blitt så komplisert de siste 10-15 år at definisjonen av student er vanskelig. Det vil også ha betydning for kvaliteten på studentdataene i NUDB. Vi har imidlertid alltid måtte forholde oss til tallet på rapporterte registrerte studenter ved lærestedene.

Når det gjelder registrering av fullførte utdanninger har det største problemet gjennomgående vært registrering av oppnådde cand. mag.- grader. Dels skyldes dette underrapportering fra flere læresteder og dels galt tidspunkt for registrering av oppnådd cand. mag.- grader. Årsaken til det er at i store deler av perioden har rapporteringstidspunktet vært identisk med tidspunktet for utsteding av vitnemål som kan ha vært lenge etter at cand. mag.-graden faktisk er oppnådd.

Handelshøgskolen BI

De mest omfattende feil og mangler i dataene som er blitt benyttet i NUDB, har bakgrunn i mangelfull rapportering fra Handelshøgskolen BI gjennom mange år og problemer med avklaringer rundt definisjoner av BIs studentmasse i forhold til våre kriterier. Det har også vært en del feil i rapportering av ulike former for fullførte utdanninger, ikke minst når det gjelder fullføringstidspunkt. Dette gjelder også siviløkonomer. Først når studentene har bedt om vitnemål er disse blitt utstedt og utdanningene er rapportert som fullført.

Dette er svært beklagelig da Handelshøgskolen BI er den største private høgskolen i landet og mangelfulle data fra dette lærestedet har betydning for kvaliteten i NUDB når det gjelder økonomisk utdanning på universitets- og høgskolenivå. Disse problemene har i større og mindre omfang eksistert i hele perioden, og til tross for mye oppklaringsarbeid har det vært vanskelig å få løst alle problemer. En del problemer skyldes også at BIs definisjoner av studenter og fullførte utdanninger/registrering av tidspunkter for fullførte utdanninger vanskelig har latt seg tilpasse SSBs krav.

Det vises for øvrig også til del 2 av dette notatet om prinsipper, definisjoner og avgrensning av massen.

8.2.2. 1970-1979

Befolkningens høyeste utdanning (BHU) 1970-1979

BHU-registeret i 1970 baserer seg på opplysninger om høyeste fullførte utdanning gitt i folke- og boligtellingsen i 1970. Feil og mangler kan skyldes at noen har gitt feil opplysninger om deres utdanning i skjemaet de har fylt ut. Noen få har heller ikke fylt ut skjemaet eller de har ikke gitt opplysninger om sin høyeste fullførte utdanning. Hvor mange som ikke ga korrekte opplysninger om sin utdanning i tellingen, er det ikke mulig å si noe om, men antagelig dreier det seg om et lite antall så kvaliteten skulle være god i alle fall når det gjelder nivå og fagfelt- de to første siffer i den seksifrede utdanningskoden. Kvaliteten på utdanningsdataene i folketellingen i 1970 betyr mye da disse dataene danner grunnlaget for alle BHU-registrene.

Det ble ikke hentet inn data om fullførte utdanninger i utlandet i perioden 1970-1979 for personer som var registrert bosatt i Norge og for personer som innvandret til Norge for første gang i perioden. Hvis vi skal lage bhu-registre for årene 1970-1979, vil det mangle data for fullførte utdanninger 1971-1973. Både utdanning fra utlandet i perioden 1971-1980 og fullførte i 1971-1973 er med i 1980, dersom de fortsatt er registrert bosatt i landet.

Igangværende utdanninger 1970-1979

Det eksisterer årgangsfiler med igangværende utdanninger fra og med 1974. Kvaliteten er gjennomgående god, men det vil alltid være enkelte mindre mangler i dataene som særlig skyldes mangelfull rapportering.

I 1976 manglet det totalt individopplysninger for 5 600 elever i Osloskolene dette året. Det manglet også 150 elever fra de daværende videregående skoler i Finnmark. I tillegg manglet det om lag 200 elever i den daværende gymnasutdanningen og 300 elever i den daværende handelsskoleutdanningen.

I 1979 ble det ikke innrapportert oppgaver over 250 elever fra Osloskolene. Det manglet oppgaver fra Troens Bevis Bibel- og Misjonsskole og Oslo Sprogskole. Ufullstendige opplysninger fra en skole i Hedmark førte til at det ble rapportert mellom 50 og 100 for få elever.

Det er ikke kjent at det var særskilte mangler i dataene fra universitetene og høgskolene i årene 1974-1979. Se for øvrig under generelt.

Fullførte utdanninger 1970-1979

Grunnskoledataene er av høy kvalitet i hele perioden 1970-1979.

Det eksisterer årgangsfiler med fullførte utdanninger fra og med skoleåret 1974/75. Kvaliteten er gjennomgående god, men det vil alltid være enkelte mindre mangler i dataene som skyldes mangelfull rapportering.

Høsten 1975 var det enkelte videregående skoler i Oslo som ikke rapporterte på individnivå. Det er derfor ikke oppgitt hvor mange som avbrøt eller ikke fullførte utdanningen i skoleåret 1975/76. Alle er blitt definert som fullført i statistikken. Det har ført til at tallet på fullførte utdanninger er litt for høyt dette året og at tallet på avbrutte/ikke fullførte utdanninger er litt for lavt.

Høsten 1976 var det 5 000 elever ved Osloskolene som fullførte en utdanning som ikke ble rapportert og som ikke inngår i dataene over fullførte utdanninger for 1976/77. I tillegg til dette var det noen som avbrøt sin utdanning dette skoleåret, men som likevel ble definert som fullført i dataene.

Det er ingen kjente mangler i rapporteringen av fullførte data fra universitetene og høgskolene i perioden 1974/75-1979/80.

8.2.3. 1980-1989

Befolkningens høyeste utdanning (BHU) 1980-1989

I den skjemabaserte 1980 - folke- og boligtellingsen ble det spurt om fullført utdanning i utlandet i perioden 1970-1980. BHU-registeret for 1980 bygger derfor på BHU-registeret for 1970, innrapporterte fullførte utdanninger i Norge årene 1970-1980 og fullførte utdanninger i utlandet årene 1970-1980. BHU-registeret for 1980 er derfor av meget god kvalitet.

Det er blitt laget BHU- årgangsfiler hvert eneste år fra og med 1985. Ved uttak fra NUDB kan det lages BHU-årgangsfiler for alle årene 1980-1989.

Jo lenger vi kommer ut på 1980-tallet jo dårligere blir kvaliteten på BHU-registeret. Det skyldes manglende opplysninger om utdanninger fullført i utlandet på 1980-tallet. Dette gjelder både for de som var registrert bosatt i Norge og som tok hele eller deler av sin utdanning i utlandet i perioden 1980/81- 1983/84 og for de som innvandret til landet etter 1980 som vi ikke hadde opplysninger om. Det foreligger altså opplysninger om fullførte utdanninger i utlandet i perioden 1984/85-1989/90 for de som var registrert bosatt i Norge. Mot slutten av 1980-tallet var det svært mange med uoppgitt utdanningsnivå for de som hadde innvandret til Norge etter 1980. Det må understrekes at utdanninger fullført i utlandet bare er kodet på tosiffernivå dvs. utdanningsnivå og fagfelt.

Igangværende utdanninger 1980-1989

Det er ingen større feil i elevdataene for videregående skoler som er registrert i perioden 1980-1989.

I 1985 ble det imidlertid rapportert alt for få studenter fra Handelshøgskolen BI. Dette året fikk vi bare rapportert de nye studentene ved lærestedet. Tallet på studenter ved lærestedet var derfor om lag 6 000 for lavt dette året. Tallet for det daværende Universitetet i Trondheim er også noe for lavt, men det er ikke mulig å fastslå et nøyaktig tall. Året etter var tallet på rapporterte studenter ved Handelshøgskolen BI 5-6 000 for høyt. Det er viktig å understreke disse betydelige manglene for disse to årene da Handelshøgskolen BI er et viktig lærested for økonomisk utdanning i Norge.

Fra 1985 er det blitt rapportert for få studenter fra Trondheim Økonomiske Høgskoles kursvirksomhet (nå integrert i Høgskolen i Sør-Trøndelag). Fram til i dag er det rapportert 2 718 for få studenter. Det er usikkert hvordan disse fordeler seg på de ulike år. Utdanningene (kursene) er av forholdsvis kort varighet, men de gir studiepoeng og skulle ha vært med i den rapporterte studentmassen.

Fullførte utdanninger 1980-1989

Det er ingen større feil i dataene for videregående skoler som er registrert i perioden 1980-1989. Mangelfull rapportering av cand.mag.-grader og mulig feil registreringstidspunkt er nok det som betyr mest på universitets- og høgskolenivå.

Fra 1985 og fram til i dag er det imidlertid blitt rapportert for få fullførte utdanninger fra Trondheim Økonomiske høgskoles kursvirksomhet (nå integrert i Høgskolen i Sør-Trøndelag). Total dreier det seg om 454 fullførte utdanninger i hele perioden, men det er usikkert hvordan de fordeler seg på de forskjellige år.

8.2.4. 1990-1998

Befolkningens høyeste utdanning (BHU) 1990-1998

I 1990/91 ble det gjennomført en undersøkelse om utdanning fullført i utlandet blant alle som var født i utlandet og som hadde innvandret til Norge etter 1980. Svarprosenten var forholdsvis høy - om lag 70 prosent - , og utdanningsopplysninger gitt i denne undersøkelsen bidro til at andelen med uoppgitt utdanning i BHU-registeret ble redusert mye igjen. Også disse utdanningene ble kodet på tosiffernivå, dvs. på nivå og fagfelt.

På samme måte som for 1980-tallet ble kvaliteten på BHU-registeret dårligere ut over på 1990-tallet. Hovedårsaken til dette var også den samme; mangelfulle data om utdanning fullført i utlandet for alle som innvandret til Norge for første gang - nå gjaldt det 1990-tallet. Da innvandringen fra enkelte land var stor dette tiåret, ble kvaliteten på BHU-registeret for befolkningen med bakgrunn i disse land redusert etter hvert. Andelen med uoppgitt steg, og det var igjen behov for en ny intervjuundersøkelse om utdanning fullført i utlandet.

Igangværende utdanninger 1990-1998

I årene 1990-1998 er det ingen større feil eller mangler i elev- og studentdataene som vi kjenner til. Dette gjelder alle data innen videregående opplæring og universiteter og høyskoler som er blitt benyttet i NUDB.

Se 1980-1989 ang. Høgskolen i Sør-Trøndelag.

Fullførte utdanninger 1990-1998

Se 1980-1989 ang. Høgskolen i Sør-Trøndelag.

Skoleåret 1989/90 ble det ikke rapportert fullførte grunnskoleutdanninger for Oslo, Buskerud, Vestfold, Telemark, Aust-Agder og Vest-Agder. Som erstatning for disse dataene ble brukt antall 15-åringer som var bosatt i de respektive fylker. Det samme gjaldt for Oslo også i skoleåret 1991/92.

Det er også en feil i eierforholdskode for fullført grunnskole 1989/90. Alle som har "uoppgitt" i skolekode har fått "privat" i eierforhold. Disse skulle hatt "uoppgitt" i eierforhold. Dette gjelder for 59 696 personer.

8.2.5. 1999-

Igangværende utdanning høst 1999 og fullførte utdanninger studieåret 1998/99 var første årganger etter en større faglig og teknologisk omlegging av den individbaserte utdanningsstatistikken. Se mer om dette i kap. 8.3.

Befolkningens høyeste utdanning (BHU) 1999-

I 1999 ble det gjennomført en ny undersøkelse om utdanning fullført i utlandet. Respondentene i denne undersøkelsen var alle med utenlandsk bakgrunn som var registrert bosatt i Norge og som sto med uoppgitt utdanning i registeret over befolkningens høyeste utdanning. De som ikke svarte på undersøkelsen i 1990/91 ble også inkludert. I tillegg ble også nordmenn som bodde i utlandet da BHU ble opprettet i 1970 og som sto med uoppgitt utdanning i BHU, tatt med i undersøkelsen i 1999. I og med denne undersøkelsen ble andelen med uoppgitt utdanning i BHU-registeret igjen mye redusert og kvaliteten derfor bedret betraktelig. Kvaliteten vil nå gjennomgående bli noe dårligere år for år på

grunn av at mange av de som har innvandret til landet de siste årene vil ha uoppgitt som høyeste fullførte utdanning da vi mangler opplysninger om deres fullførte utdanning i utlandet.

Igangværende utdanninger 1999-

Se 1980-1989 ang. Høgskolen i Sør-Trøndelag

For igangværende 1.10.2003 ble AMO-kurs med en varighet på under 300 timer tatt med i dataene. Dette omfatter anslagsvis 5000 personer.

Grunnlaget for statistikk for videregående opplæring er "elevkurslinjer" i VIGO, som er fordelt på skoleår. Etter innføring av rett til videregående opplæring for voksne i 2000 ble det observert en økning i antall av voksne i innsamlingen og økt antall dubletter hvor voksne var registrert på flere kursnivåer. Det ble imidlertid ikke iverksatt tiltak i forhold til voksne før i 2003. Det var da kjent at mange voksne ikke var registrert i VIGO i det hele tatt, alternativt at informasjonen om kurs ikke var nøyaktig i og med at voksne følger særskilte løp, f.eks. følger flere kurstrinn samtidig. Voksne som søker videregående opplæring iht. rett til videregående opplæring som voksne er i hovedsak registrert med ønsket sluttkompetanse, uavhengig av hvilke kurs de eventuelt faktisk følger.

I 2003 ble innført en variabel som registrerte omfanget av personer som var realkompetansevurdert forut for skoleåret. Denne gruppen kan skilles ut i statistikken (~~elevtype=R~~)-(elevstatus=V). Gruppen har i noen grad også vært rapportert tidligere, omfanget av økning i rapportering som følge av en egen variabel er ikke kjent. I statistikken førte rapporteringen til en overvurdering av personer i fagopplæring som ikke var rapportert som lærlinger. I tallene for igangværende 1.10.2003 er personer registrert med sluttkompetanse fagbrev og ikke samtidig registrert som lærlinger plassert i VKI, dette utgjør i underkant av 2.000 elever.

I 2004 ble det etablert en ny innsamlingsordning (egen fil) for voksne som ikke er relatert til elevkurslinjer i VIGO, men til ønsket sluttkompetanse og oppfølging av denne. Innsamling fra fylkeskommunene av voksne på egen fil er ikke lagt inn i NUDB. Elevstatus V forekommer også på elevfilene, de inngår da i den ordinære elevstatistikken og legges inn i NUDB på linje med elever.

Pr. rapporteringsåret 2005 er voksne (definert ved elevstatus V) ikke fjernet fra ordinær elevstatistikk. Hovedgrunnen er manglende kunnskap om i hvilken grad gruppen er representert i elevtallene allerede fra skoleåret 2000/2001.

I spesielle tilfeller kan elever gå rett fra grunnskole ut i lære uten å gå veien om skole. Disse er plassert på NUS-kode for lærlinger/fagprøver. Det betyr at de er plassert på tredje året (andre og tredje året for særløp) under hele utdanningen.

Fullførte utdanninger 1999-

Det er ingen større kjente feil eller mangler i denne perioden verken for videregående utdanning eller for høyere utdanning.

Se 1980-1989 ang. Høgskolen i Sør-Trøndelag.

Registrering av avlagte studiepoeng i perioden 1999-2002.

Det ble foretatt en endring i registreringen av fullførte utdanninger fra og med skoleåret 1998/99. Omleggingen innebar at bare grader registreres som fullført utdanning. Utdanninger som ikke er definert som grader registreres bare med avlagte studiepoeng. Dette omfatter stort sett lavere grads utdanninger og videreutdanninger. I dataene (og i utdanningsstandarden) har disse utdanningene kodetype=1.

I årene 1999-2002 rapporterte ikke de høgskolene som hadde annet datasystem enn FS avlagte studiepoeng på emnenivå. De rapporterte i prinsippet som før, dvs. at tilleggsutdanninger ble rapportert som avsluttet med utfall=8 fullført eller med utfall=2 ikke fullført. Disse avsluttarecordene hadde likevel delvis påført avlagte studiepoeng. En del recorder ble meldt som fullført(utfall=8), men hadde likevel ikke fått påført antall avlagte studiepoeng. Det er grunn til å tro at disse burde hatt avlagte studiepoeng. I NUDB er disse recordene uten studiepoeng ikke skilt fra recorder med utfall=2(ikke fullført). Begge er registrert med en record, med 0 avlagte studiepoeng.

8.3. Prinsipper, definisjoner og avgrensing av massen

8.3.1. Generelt

Grunnlagsfilene for NUDB er alle årgangsfiler med igangværende og fullførte utdanninger fra og med 1974 og foreløpig fram til og med 2005. I tillegg kommer alle registrene med befolkningens høyeste utdanning (BHU). Over et så langt tidsrom vil det nødvendigvis måtte bli en del endringer i definisjoner av masse og variabler på årgangsfilene. Ulike endringer av utdanningsstrukturen er nok det som har betydd mest for endringene. Det vil ikke her bli redegjort for alle disse, men vi skal gi en kort oversikt over de viktigste endringene som er blitt gjort i perioden og som kan være av betydning for brukere av NUDB- data. Gjennomgangen er kronologisk.

Helt sentralt i alle NUDB-data er den sekssifrede utdanningskoden i NUS 2000. I NUDB er det kun NUS 2000- koder som brukes, dvs. at alle utdanningskoder som bygger på gamle utdanningsstandarder er blitt omkodet til NUS 2000. Når det skal lages tilbakegående tall med grunnlag i NUDB er det derfor NUS 2000 - kodene som vil være grunnlaget. Det er mye vanskeligere å bruke en sentral variabel som skoleslag hvis man skal bruke data fra mange årganger da det har vært mange endringer av inndelingen i skoleslag i perioden. Og mange utdanninger er blitt flyttet fra videregående nivå til universitets- og høgskolenivå. Det må understrekes at på grunn av omkodinger vil f.eks. alle utdanninger som i dag er høgskoleutdanninger også være det på 1970- tallet i NUDB. F.eks. vil en helsefagutdanning som på 1970- tallet var en videregående utdanning og som fra og med 1981 ble omgjort til en høgskoleutdanning, ha den samme utdanningskoden hver årgang i NUDB. Fra begynnelsen av 1980 og fram til i dag har det imidlertid vært forholdsvis få endringer av nivåplassering av utdanninger, men flere endringer av strukturer innenfor videregående utdanning og universitets- og høgskoleutdanning. Noen strukturendringer skjedde etter endringer av eksisterende lover eller etter at det ble vedtatt nye lover.

I hele perioden har det vært søkt å avgrense massen av elever og studenter som inngår i dataene på videregående nivå. Det er satt et skille ved 300 undervisningstimer per år, dvs. at alle som deltar i en utdanning med 300 undervisningstimer eller mer per år skal inngå i dataene, mens de med færre enn 300 undervisningstimer per år holdes utenfor. Fra og med 2001 ble statistikken for videregående skoler under Opplæringsloven utvidet til også å omfatte utdanninger med færre enn 300 timer. Avgrensing av massen gjaldt også i prinsippet for universitets- og høgskoleutdanninger fram til begynnelsen av 1990- tallet, men regelen hadde ingen verdi på dette nivået da tallet på undervisningstimer var lite relevant for universitetene og høgskolene. Det ble i 1992 satt en vekttallsgrense på 5 vekttall i semesteret for høgskolesektoren, og denne grensen gjaldt i prinsippet fram til omleggingen av den individbaserte utdanningsstatistikken på slutten av 1990-tallet. I universitetssektoren (universiteter og vitenskapelige høgskoler) har det aldri vært forsøkt å sette en vekttallsgrense.

For igangværende utdanning inngår lærlingene fra og med 1.10.1989. Tidligere var elever ved lærlingskoler med i massen. Fra og med 1989 er det ikke registrert elever i lærlingskoler. Når elevtall før og etter 1989 skal sammenliknes, må lærlingetall og elever i lærlingskole holdes utenfor.

Studenter i utlandet som har hatt lån i Statens lånekasse er lagt inn i dataene fra 1.10.1986. Tidligere har dataene fra Statens lånekasse bare blitt benyttet til å oppdatere befolkningens utdanningsnivå.

Når det gjelder fullførte utdanninger er det aller viktigst å understreke mangelen på enhetlige definisjoner i registreringen av fullførte utdanninger ved universitetene på 1970- 1980- og 1990- tallet. På de matematisk- naturvitenskapelige fakultetene ble det ikke registrert fullførte utdanninger før oppnådd cand.mag.- grad, mens det på de andre fakultetene ble registrert fullførte grunnfag, mellomfag, årsheter og påbygging på årsheter. Dette er en generell svakhet i registreringen av fullførte utdanninger ved universitetene. Ved høyskolene ble alle fullførte utdanninger registrert. Ved fullføring av lavere og høyere universitetsgrad bestående av flere utdanninger (fag) ble ikke det siste faget i graden registrert som fullført fram til omleggingen av utdanningsstatistikken i 1999. Graden ble registrert. Det må understrekes at det alltid har vært en noe mangelfull rapportering av cand.mag.-grader (se kap. om feil og mangler).

I hele perioden er alle fullførte utdanninger på videregående nivå av en varighet på 300 undervisningstimer eller mer blitt registrert. Det viktigste ved registreringer av fullførte utdanninger på videregående nivå er alle de endringer i regler for utsteding av vitnemål som har funnet sted de siste 25-30 år. Dette har ført til mange forskjellige koder for fullført utdanning i ulike perioder.

Registeret over befolkningens høyeste utdanning (BHU) ble før omleggingen av utdanningsstatistikken oppjustert med alle fullførte utdanninger (inkl. ex. phil). hvis utdanningen var på et høyere eller tilsvarende nivå som utdanningen som tidligere var registrert som høyeste fullførte utdanning for personen. Dette slo noe skjevt ut på universitets- og høyskolenivå på grunn av den mangelfulle registreringen av fullførte utdanninger ved mat. nat.- fakultetene ved universitetene (se ovenfor). Fra og med 1999- oppjusteringen er bhu- registeret først oppjustert på uh-nivå når en 20-vektallsgruppe er fullført. Etter et større utredningsarbeid ble nivådefinisjonene i bhu endret i 2006.

8.3.2. 1970-1979

Befolkningens høyeste utdanning (BHU) 1970-1979

BHU- registeret fra 1970 er det første BHU- registeret. Grunnlaget for registeret er dataene som ble hentet inn i folke- og boligtellingsen i 1970. Registeret omfatter alle personer registrert bosatt i Norge 1.11.1970 som per 31.12. i 1970 var 16 år eller mer samt yngre personer som hadde fullført en utdanning.

Grunnleggende fagutdanning før 1972/73 som er basert på 9-årig grunnskole, er plassert på klassetrinn 10 (førstesiffer 3 i utdanningskoden). Det ble ikke laget noe nytt BHU-register før etter folke- og boligtellingsen i 1980.

Igangværende utdanninger 1974-1979

Videregående skoler

De videregående skolene ble i 1974 og 1975 gruppert etter Standard for næringsgruppering 72. Videregående utdanning omfattet de fleste av elevene som var å finne på klassetrinn 10-12.

Følgende skoleslag ble brukt på årgangsfilene 1974 og 1975:

- Folkehøgskoler
- Realskoler
- Gymnas
- Andre allmennskoler
- Fag- og yrkesskoler
- Fagskoler for landbruk og fiske
- Fag- og yrkesskoler for håndverk og industri

- Fag- og yrkesskoler for transport og kommunikasjon
- Skoler for administrasjon og økonomi
- Skoler for helsestell og sosialt arbeid
- Fag- og yrkesskoler ellers

I 1975 vedtok Stortinget en ny lov om videregående utdanning.

Fra og med 1976 ble Standard for næringsgruppering 78 benyttet på årgangsfilene.

Hovedgruppene var som følger:

- Skoler under lov om videregående opplæring (Inkl. tidligere gymnas og fag- og yrkesskoler)
- Folkehøgskoler
- Skoler for helsestell og sosialt arbeid
- Fagskoler for landbruket
- Videregående skoler ellers

På slutten av 1970- tallet ble en del estetiske skoler omgruppert fra videregående skoler til høgskoler.

Universiteter og høgskoler

Omfang av universiteter og høgskoler var alle som etter Standard for næringsgruppering 72 og 78 ble definert som universitet eller høgskole.

Årene 1974 og 1975 ble Standard for næringsgruppering 72 benyttet. Inndelingen i skoleslag var som følger:

- Universiteter
- Distriktshøgskoler
- Lærerhøgskoler
- Tekniske skoler
- Sosialskoler
- Skoler for videreutdanning i sykepleie
- Militære høgskoler
- Andre høgskoler

Fra og med 1976 ble Standard for næringsgruppering 78 benyttet. Inndelingen i skoleslag ble da som følger:

- Universiteter og vitenskapelige høgskoler
- Distriktshøgskoler
- Pedagogiske høgskoler
- Ingeniørhøgskoler
- Sosialhøgskoler
- Helsefaghøgskoler
- Militære høgskoler
- Andre høgskoler

De fleste universitets- og høgskolestudentene var å finne på klassetrinn 13-18.

Det er viktig å understreke at på 1970-tallet hørte bl.a. sykepleierskolene inn under videregående opplæring. Det samme gjorde også de maritime skolene.

En omgruppering av skoler på slutten av 1970- tallet førte som tidligere nevnt til at en rekke estetiske skoler ble omgjort til høgskoler.

Fullførte utdanninger 1970-1979

Videregående skoler

Det var ingen vesentlige endringer i definisjoner av hva som skulle inngå i massen for fullførte utdanninger på 1970- tallet, men før 1976 var det kun de som hadde bestått som fikk vitnemål. Etter dette år fikk også de som hadde strøket i et eller flere fag vitnemål.

Universiteter og høyskoler

Det var heller ingen vesentlige endringer i definisjoner av hva som skulle inngå i massen for universitetene og høyskolene for fullførte utdanninger på 1970- tallet.

8.3.3. 1980-1989

Befolkningens høyeste utdanning (BHU) 1980-1989

Det eksisterer BHU-register for 1980 og for hvert av årene 1985-1989.

Oppdateringen foregikk på den samme måten hele tiden, og det var ingen endringer av definisjoner eller av prinsipper for oppdateringen av registeret i denne perioden. Grunnlaget er BHU-registeret for 1970 og alle fullførte utdanninger i Norge fra og med dette året. Grunnlaget er også opplysninger om fullførte utdanninger i utlandet som ble gitt i folke- og boligtellingsen i 1980 og opplysninger om fullførte utdanninger i utlandet fra og med 1982/83 for alle registrert bosatt i Norge. De sistnevnte dataene er innhentet fra Lånekassen.

Igangværende utdanninger 1980-1989

Videregående skoler

På hele 1980- tallet var det Standard for næringsgruppering 1978 som ble brukt.

Det var derfor de samme hovedgrupper av videregående skoler:

- Skoler under lov om videregående opplæring
- Folkehøgskoler
- Skoler for helsestell og sosialt arbeid
- Fagskoler for landbruket
- Videregående skoler ellers

I 1981 ble en rekke skoler i gruppen "Skoler for helsestell og sosialt arbeid" overført til universitets- og høyskolesektoren. Viktigst var sykepleierskolene.

Også Statens operaskole og Statens teaterskole og Ballettinstituttet ble overført fra videregående skoler til universiteter og høyskoler.

I 1982 ble maritime skoler og avdelinger ved Handelshøyskolen BI overført fra videregående skoler til universiteter og høyskoler.

Elever i videregående utdanning var på 1980- tallet hovedsakelig å finne på klassetrinn 11-13.

Lærlinger inngår på filene med igangværende utdanninger for første gang i 1989. Tidligere var elever ved lærlingskoler med i massen. Fra og med 1989 er det ikke registrert elever i lærlingskoler. Når elevtall før og etter 1989 skal sammenlignes, må lærlingetall holdes utenfor.

Universiteter og høgschooler

Standard for næringsgruppering 78 ble brukt på hele 1980- tallet, dvs. følgende inndeling i skoleslag:

- Universiteter og vitenskapelige høgschooler
- Distriktshøgschooler
- Pedagogiske høgschooler
- Ingeniørhøgschooler
- Sosialhøgschooler
- Helsefaghøgschooler
- Militære høgschooler
- Andre høgschooler

I 1981 ble mange helsefagskoler omgjort til høgschooler - sykepleierskoler, vernepleierskoler og radiografskoler. Det samme gjaldt for Statens operaskole og Statens teaterskole.

I 1982 ble maritime skoler og avdelinger ved Handelshøgschoolen BI overført fra videregående skoler til universiteter og høgschooler.

Universitets- og høgschoolestudentene var i hovedsak å finne på klassetrinn 14 og over.

Studenter i utlandet som har hatt lån i Statens lånekasse er lagt inn i dataene fra 1.10.1986. Tidligere har dataene fra Statens lånekasse bare blitt benyttet til å oppdatere befolkningens utdanningsnivå.

Fullførte utdanninger 1980-1989

Videregående skoler

Det var på 1980- tallet ingen større endringer i hva som inngikk i massen for fullførte utdanninger.

Universiteter og høgschooler

Det var på 1980- tallet ingen større endringer i hva som inngikk i massen for fullførte utdanninger.

8.3.4. 1990-1998

Befolkningens høyeste utdanning (BHU) 1990-1998

Det er blitt laget BHU-registre for hvert år i denne perioden.

I 1991 ble det gjennomført en intervjuundersøkelse om utdanning fullført i utlandet. Alle som innvandret til Norge for første gang i perioden 1981-1990 ble spurt om deres høyeste fullførte utdanning i utlandet. BHU-registeret ble for øvrig oppjustert etter samme prinsipper i hele denne perioden.

Se også 3.5.

Igangværende utdanninger 1990-1998

Videregående skoler

Standard for næringsgruppering 78 ble benyttet til og med 1992

Hovedgrupper:

- Skoler under lov om videregående opplæring
- Folkehøgschooler
- Skoler for helsestell og sosialt arbeid
- Fagskoler for landbruket
- Videregående skoler ellers

Ny standard for næringsgruppering 94 ble tatt i bruk fra og med 1993-årgangen.

- Undervisning i allmennfag
- Undervisning i tekniske og andre yrkesrettede fag (inkl. lærlinger)
- Folkehøgskoleundervisning
- Arbeidsmarkedskurs

I 1994 ble det gjennomført en ny og viktig reform i videregående utdanning - Reform 94. Alle i alderen 16-19 år fikk rett til treårig (fireårig) videregående opplæring som skulle føre fram til studiekompetanse eller yrkeskompetanse/fagbrev. Denne retten ble noen år seinere utvidet til også å gjelde voksne.

Universiteter og høyskoler

Til og med 1992 ble universitetene og høyskolene inndelt etter Standard for næringsgruppering 78 med denne grupperingen:

- Universiteter og vitenskapelige høyskoler
- Distriktshøyskoler
- Pedagogiske høyskoler
- Ingeniørhøyskoler
- Sosialhøyskoler
- Helsefaghøyskoler
- Militære høyskoler
- Andre høyskoler

Fra og med 1993 ble Standard for næringsgruppering 1994 benyttet. Denne grupperingen blir benyttet:

- Undervisning ved universiteter
- Undervisning ved statlige høyskoler
- Undervisning ved militære høyskoler
- Undervisning ved andre høyskoler

Innføringen av ny næringsstandard falt sammen i tid med at det ble gjennomført en større reform i universitets- og høyskolesektoren. De gamle regionale høyskolene som ingeniørhøyskoler, pedagogiske høyskoler og helsefaghøyskoler ble slått sammen til større regionale statlige høyskoler.

Fra og med 1992 ble det som tidligere nevnt satt en vekttallsgrense for hvilke studenter som skulle inngå i høyskoledataene. Kun de studenter som var registrert på studieprogrammer som totalt omfattet 5 vekttall eller mer i semesteret, skulle inngå i massen. Dette ble gjort fordi 300-timersregelen ikke hadde noen relevans for universitetene og høyskolene. 5-vektallsgrensen gjaldt imidlertid ikke for universitetssektoren da det ikke var mulig å få til en slik avgrensning for denne. Hensikten var ikke å foreta en ny avgrensning av massen, men heller å presisere bedre hva dataene fra høyskolene skulle omfatte.

Fullførte utdanninger 1990-1998

Videregående skoler

Det ble ikke gjort noen vesentlige endringer i prinsipper og definisjoner når det gjelder fullførte utdanninger i perioden. Fra og med 1994 ble det ikke utstedt vitnemål til de som strøk i ett eller flere fag, men kun delkompetansebevis.

Universiteter og høgschooler

Det ble heller ikke for universitets- og høgschoolsektoren gjort noen vesentlige endringer i prinsipper og definisjoner når det gjelder fullførte utdanninger i perioden.

Studieåret 1994/95 var det svært få som fullførte allmennlærerutdanning. Det henger sammen med at allmennlærerutdanning ble utvidet til fire år for de som begynte på en slik utdanning i studieåret 1991/92. På grunn av utvidelsen i antall studieår blir det derfor naturlig nok svært få som slutter det ene året.

8.3.5. 1999-

Befolkningens høyeste utdanning (BHU og BU) 1999-

Det er blitt foretatt en endring av oppjusteringen av bhu for de som har fullført en videregående utdanning etter Reform94. Det er nå bare de som har registrert "bestått" i fullførkode etter å ha fullført en videregående utdanning som blir oppjustert i bhu. Dette gjelder enten det er en fullført GK-, VKI- eller VK II- utdanning. Endringene har ført til at noen som tidligere at noen som tidligere stod i bhu med en videregående utdanning som sin høyeste fullførte utdanning nå er blitt nedjustert til grunnskolenivå.

I 1999 ble det igjen gjennomført en intervjuundersøkelse der alle som hadde innvandret til Norge for første gang etter 1990 og alle som ikke hadde svart på den forrige undersøkelsen i 1991 ble spurt om deres høyeste fullførte utdanning fra utlandet. Svarprosenten var over 80 prosent, og andelen med uoppgitt i BHU- registeret sank betraktelig. Samtidig ble prinsippene for oppjustering av BHU endret. Disse endringene hadde stor betydning for universitets- og høgschoolnivå. Disse gjaldt fram til 2006 da nye nivådefinisjoner ble innført. Hovedkriteriene for å fastslå hvilken utdanning som er den høyeste fullførte er nå som følger:

Utdanningen med høyest nivå (1.siffer i utdanningskoden) velges

**Hvis nivå ikke skiller, velges utdanningen med høyest klassetrinn*

**Hvis klassetrinn heller ikke skiller, velges den utdanningen hvor ikke 2.siffer, dvs. fagfelt, i utdanningskoden = 0 (allment fagfelt)*

**Hvis heller ikke fagfelt skiller, velges utdanningskoden med den nyeste avslutningsdatoen*

**Hvis heller ikke avslutningsdato skiller, velges den første utdanningen, dvs. utdanningen som allerede er registrert i BHU*

Studenter som tok utdanninger som ikke var profesjonsutdanninger ved et universitet eller en høgschool, ble registrert med en ny utdanning på BHU-filen hver gang de hadde fullført en 20-vektallsgruppe (ett års studium), ev. en 10-vektallsgruppe (ett semesters studium) når det gjaldt en videre- eller etterutdanning. Det må understrekes at den utdanningskoden som en person ble registrert med på BHU-filen, etter at vedkommende hadde fullført en 20-vektallsgruppe, var den sist fullførte emnegruppen, uavhengig av hvor mange vektall det var i denne emnegruppen i forhold til de andre emnegruppene som utgjorde 20-vektallsgruppen. Enkelte personer som står registrert med en utdanning på nivå 6 i NUS2000 vil derfor ikke nødvendigvis ha korrekt 2.siffer, dvs. fagfelt, i utdanningskoden.

Følgende nivådefinisjoner gjaldt i bhu fram til endringer av nivådefinisjoner i 2006 (se nedenfor).

grunnskolenivå

-alle som hadde fullført en grunnskoleutdanning

NUS2000- nivå = 1 og 2

ISCED97 - nivå = 1 og 2

videregående nivå

-alle som hadde fullført en eller annen videregående utdanning uansett utdanningens varighet og alle som hadde fullført påbygging til videregående utdanning

NUS2000- nivå = 3, 4 og 5

ISCED97 - nivå = 3 og 4

universitets- og høghskolenivå - lavere grad

-alle som har fullført en universitets- eller høghskoleutdanning av en varighet på inntil fire år

-fra 1999 ble man først definert på uh- nivå etter å ha fullført minst 20 vekttall (60 studiepoeng) i uh-systemet. Dette gjaldt imidlertid ikke for fullførte profesjonsutdanninger som måtte være fullført i sin helhet før oppjustering til uh- nivå. De første årene etter omleggingen var rapporteringen av antall avlagte studiepoeng mangelfull. I enkelte prosjekt kan det derfor være mer korrekt å ta med personer som står registrert med et lavere antall studiepoeng enn 60 på uh- nivå. Ved å ta med disse antar vi at vil en komme nærmere de korrekte tallene for f.eks. tilleggsutdanninger. Vi må likevel kunne anta at det har vært en underrapportering av tilleggsutdanninger avlagt i årene 1999-2002. Fra 2003 er det grunn til å anta at rapporteringen har vært korrekt.

NUS2000- nivå = 6 (minst 20 vekttall (60 studiepoeng) fra 1999)

ISCED97 - nivå = 5 (kltr 14-17)

universitets- og høghskolenivå - høyere grad

-alle som hadde fullført en universitets- eller høghskoleutdanning av en varighet på mer enn fire år og alle som hadde fullført en forskerutdanning

NUS2000- nivå = 7 og 8

ISCED97 - nivå = 5 (kltr 18-19) og 6

Nye definisjoner av utdanningsnivå:

Etter et ganske omfattende utredningsarbeid ble det i 2006 innført nye nivådefinisjoner i bhu. Hovedårsakene til at endringene ble gjort var å få norske nivådefinisjoner mer i samsvar med internasjonale retningslinjer. I NUDB ligger både den gamle(bhu) og den nye(bu) definisjonen av utdanningsnivå. Den nye bu-inndelingen er etablert for alle år, mens den gamle bhu-inndelingen ikke vil bli videreført etter 2005.

På grunnlag av de siste vurderinger er definisjonene av de ulike utdanningsnivåer i bu nå som følger:

grunnskolenivå

- alle som har fullført en grunnskoleutdanning

NUS2000- nivå = 1 og 2

ISCED97- nivå = 1 og 2

- alle som har fullført en videregående utdanning av en varighet på mindre enn to år i fra og med skoleåret 1975/76 og til og med skoleåret 1994/95, dvs. hovedsakelig utdanninger på grunnkursnivå.

NUS2000- nivå = 3 (kltr 11)

ISCED97- nivå = 3 (kltr 11)

- alle som har fullført en videregående utdanning av en varighet på mindre enn tre år fra og med skoleåret 1995/96, dvs. hovedsakelig utdanninger på grunnkursnivå og utdanninger på videregående kurs I- nivå.

NUS2000- nivå = 3

ISCED97- nivå = 3 (kltr 11 og 12)

videregående nivå

- alle som har fullført videregående utdanning uansett utdanningens lengde til og med skoleåret 1974/75

NUS2000- nivå = 3 og 4

ISCED97- nivå = 3

- alle som har fullført videregående utdanning av en varighet på to år eller mer fra og med skoleåret 1975/76 og til og med skoleåret 1994/95 dvs. hovedsakelig videregående kurs I- og videregående kurs II- utdanninger.

NUS2000- nivå = 3 (kltr 12) og 4

ISCED97- nivå = 3 og (kltr 12 og 13)

- alle som har fullført en videregående utdanning av en varighet på tre år eller mer fra og med skoleåret 1995/96, dvs. hovedsakelig videregående kurs II- utdanninger.

NUS2000- nivå = 4

ISCED97- nivå = 3 (kltr 13)

- alle som har fullført en utdanning på nivået mellom videregående nivå og universitets- og høgskolenivå, dvs. påbyggingsnivået til videregående nivå uansett periode

NUS2000- nivå = 5

ISCED97- nivå = 4

- alle som har fullført færre enn 120 studiepoeng i universitets- og høgskolesystemet fra og med 1998/99

NUS2000- nivå = 5 (færre enn 120 studiepoeng)

ISCED97- nivå = 4 (færre enn 120 studiepoeng)

universitets- og høgskolenivå - lavere grad

- alle som har fullført en universitets- og høgskoleutdanning av en varighet på inntil fire år fram til 1997/98

NUS2000- nivå = 6

ISCED97- nivå = 5 (kltr<18)

- alle som har fullført 120 studiepoeng eller mer i universitets- og høgskolesystemet fra og med 1998/99, men som ikke har fullført en høyere grads- utdanning

NUS2000- nivå = 6 (120 studiepoeng eller mer)

ISCED97- nivå = 5 (120 studiepoeng eller mer) (kltr<18)

universitets- og høgskoleutdanning - høyere grad

- alle som har fullført en universitets- og høgskoleutdanning av høyere grad. De fleste av disse utdanningene vil være av en varighet på 5 år eller mer.

NUS2000 - nivå = 7

ISCED97- nivå = 5 (kltr>17)

- alle som har fullført en forskerutdanning

NUS2000-nivå = 8

ISCED97- nivå = 6

Igangværende utdanninger 1999-

videregående utdanning

Fra og med 2001 ble videregående igangværende under loven, endret til også å omfatte elever som deltok med mindre enn 300 timer.

Fra 2004 er personer som går i tilbud som er spesielt tilrettelagt for voksne integrert i ordinære kurs ved videregående skoler. I andre tilfeller følger de egne tilrettelagte opplæringstilbud, for eksempel komprimerte opplæringsløp, eller følge tilbud organisert av andre tilbydere enn fylkeskommunene. I 2004 ble voksne, som i hovedsak er integrert i ordinære kurs eller egne klasser på en videregående skole, tatt med. I publisering for 2003 var også delmengder av voksne i andre undervisningstilbud spesielt tilrettelagt for voksne inkludert i elevtallene.

universitetet og høyskoler

Fra og med 2002- årgangen er det blitt innhentet opplysninger om alle registrerte studenter ved universiteter og høyskoler. Vekttallsgrensen på fem vekttall ved høyskolene ble avvirket.

Kvalitetsreformen i høyere utdanning som bl.a. har resultert i et nytt gradssystem -bachelor og master-, har gjort at det har vært nødvendig å lage mange nye NUS 2000-koder i utdanningsstandarden etter hvert som nye utdanninger er blitt etablert.

Fra 1999 ble det innført et skille mellom ordinær utdanning og fjernundervisningsstudenter, men her er kvaliteten ikke så god.

Fullførte utdanninger 1999-

videregående utdanning

Fra og med avslutta i skoleåret 2000/2001 er dataene supplert med fullføringer fra Nasjonal vitnemålsdatabase(NVB) og Helsepersonellregisteret(HPR). Dette fører til endring i utfall, mer detaljert NUS-kode eller tilganger.

universiteter og høyskoler

Fra og med studieåret 1998/99 er studiepoeng/vektall med for hver fullført emneeksamen; fra studieåret 2001/2002 studiepoeng. Dette er utdanninger av en varighet på mindre enn to år. Disse inngår i vekttalls/studiepoengsstatistikken. Vekttallene eller studiepoengene hentes fra lærestedene.

Fullførte utdanninger omfatter fra og med studieåret 1998/99 kun utdanninger med en varighet på to år eller mer (grader) inkl. praktisk pedagogisk utdanning. Vekttallene eller studiepoengene hentes fra Standard for utdanningsgruppering.

Etter hvert som studentene avlegger henholdsvis bachelor- og mastergrader vil mange av de nye utdanningskodene for bachelor og master også bli brukt for fullførte utdanninger. Det er litt vanskelig å sammenligne de nye gradene direkte med de gamle gradene.

Fra og med 2000/2001 er dataene supplert med fullføringer fra Helsepersonellregisteret(HPR).

Vedlegg

PROSJEKTSKRIV

van/- Oppr	Adm. dir.		
Fra:	Seksjonsleder Elisabetta Vassenden	Seksjon 320	Dato 2.11.00

Prosjektnavn Nasjonal utdanningsdatabase (erstatte tidligere prosjektskriv Omlegging av Befolkningens høyeste utdanning)	Produktnr. 0330
Oppdragsgiver SSB (med ev. bidrag fra KUF for å framskynde arbeidet.)	
Ansvarlig (utførende) enhet Seksjon for befolknings- og utdanningsstatistikk	
Prosjektleder Torill Vangen	

Se Statistisk sentralbyrås håndbøker 58 «Håndbok i prosjektstyring»

Bakgrunn Her beskrives kort hvorfor/hvordan prosjektet har oppstått.

Omlegging av utdanningsstatistikken har vi valgt å dele i tre delprosjekter, to delvis overlappende bearbeidingsprosjekter og et prosjekt for organisering av ferdig bearbeidede data:

Dette prosjektskrivet, UTDANNINGSDATABASEN, omfatter organiseringen av ferdig bearbeidede data.

Foruten årgangsfilene over igangværende og avsluttet utdanning er Statistisk sentralbyrås register for befolkningens høyeste utdanning (BHU) og spesielle datafiler over gjennomstrømningen i utdanningssystemet hovedsluttproduktet i dagens utdanningsstatistikk.

Informasjon om BHU-opplysninger ble innhentet på skjema for siste gang ved Folke- og bolig telling 1970 (FoB70). I forbindelse med FoB80 ble BHU oppdatert til 1.11.1980 med utgangspunkt i FoB70-dataene og ved hjelp av årgangsfilene over utdanning fullført i Norge i perioden, samt skjemainnhenting av utdanning fullført i utlandet i perioden. Fra og med 1985 er BHU-registeret oppdatert årlig med egne årgangsfiler som sluttprodukt. Disse filene er ajourført for all fullført utdanning i Norge, samt med data for utdanning fullført i utlandet med støtte fra Statens lånekasse for utdanning fra og med 1985. I forbindelse med FoB90 ble det gjennomført en egen spesialundersøkelse for utdanning fullført i utlandet blant innvandrere. Disse ble lagt inn på årgangsfilen for 1991. En tilsvarende undersøkelse er gjennomført i forbindelse med FoB2001.

Dagens register har visse svakheter som vi i ønsker å forbedre. Vi ønsker å utvikle et statistikkssystem som i større grad enn i dag, enkelt gir muligheter til å gjennomføre bredere og større analyser av all den informasjonen utdanningsstatistikkfilene inneholder, spesielt gjennomstrømningsanalyser. Statistikkssystemet skal være grunnlaget for å kunne lage gode, løpende utdanningsindikatorer, spesielt om gjennomstrømning.

Formål Forslag til målsettinger for prosjektarbeidet.

Utdanningsdatabasen skal organisere det individbaserte utdanningsstatistikkgrunnlaget slik at det er mulig på en enkel måte å utnytte de opplysningene som ligger i vårt meget omfattende materiale. Basen skal på en enkel måte:

1. Gi svar på befolkningens høyeste fullførte utdanning, samtidig som den skal kunne gi oversikter over all fullført utdanning, for et hvilket som helst år.
2. Gi svar på hvor mange som har fullført de enkelte utdanningskodene i NUS-standarden uavhengig av om utdanningen er den høyeste eller siste fullførte utdanning, per et hvilket som helst år.
3. Kunne brukes til analyser av gjennomstrømninger i utdanningssystemet, deriblant opplysninger om avbrudd og fullføringsprosjenter, bevegelsesmønster i utdanningssystemet både fremover etter en endt utdanning, og bakover i tid forut for en endt utdanning. Analysen skal kunne gjelde på et hvilket som helst tidspunkt og for en hvilken som helst periode.
4. Tilrettelegge dataene slik at muligheten for sammenliknbarhet over tid bedres.
5. Lage gode utdanningsindikatorer og utvide den løpende statistikken

Faglige problemstillinger Vektlegg kort faglige forutsetninger og problemstillinger. Dette punktet utformes i detalj senere, i forbindelse med utforming av milepælplanen.

IT-faglige

Utdanningsdatabasen skal være en forløpsdatabase i Oracle. En tilsvarende modell er allerede i bruk ved FD-Trygd-basen i SSB. I motsetning til utdanningsbasen omfatter ikke FD-Trygd-basen tabelluttaksverktøy. Prosjektet Utdanningsbase er delt i to ved del 1 omfatter tilrettelegging av data i basen og system for uttak av datafiler som kan viderebearbeides til tabeller, mens del 2 omfatter etablering av et system for uttak av tabeller fra basen.

Utdanningsfaglige/statistikkfaglige

1. Sikre sammenliknbarheten med historisk datamateriale
2. Finnes det utdanninger som krever spesialbehandling (f.eks ved fullføring i utlandet, profesjonsutdanninger osv.)
3. Sikre at det blir tatt hensyn til endrede fødselsnummer.
4. Hva skal karakteriseres som avbrudd, fullført, bestått osv.
5. Gå igjennom utvalgskriteriene for hva som er høyeste fullførte utdanning.
6. Utvikle gode indikatorer for gjennomstrømning. Definere gjennomstrømningsvariablene og utvalgskriteriene for disse.
7. Gjøre rede for de avvik som vil finnes i modellen. Hva vil de feilmarginene og manglene vi godtar i bearbeidingsystemene for igangværende og avsluttet utdanning føre til i et utdanningsforløp.

Interessenter og deltakere Hvem er oppdragsgiver, og hvem er prosjektleder. Hvem er øvrige interessenter i eller utenfor Statistisk sentralbyrå. Hvilke fagområder eller organisatoriske enheter må bidra.

Prosjektet er i utgangspunktet et omleggingsprosjekt internt på seksjon for befolknings- og utdanningsstatistikk, men de fleste seksjoner i SSB som arbeider med individbasert statistikk vil

være interessenter i prosjektet. Statistikkgrunnlagene fra utdanningsstatistikken benyttes i dag blant annet av s260, s310, s330, s340, s350, s420 og s520. Det vil også finnes eksterne interessenter i form av forskere og andre som får utdrag av dagens filer for analyse og forskningsoppgaver.

Prosjektgruppen for basen består av Torill Vangen (prosjektleder), Lise Styrk Hansen og Arild Rognan fra fagsiden. IT- siden ledes av Johnny Johansen og består videre av Anne Marie Holseter og en person til fra s303. Fra Gruppe for metoder vil Li-Chun Zhang bidra med 300 timeverk.

Følgende utgjør styringsgruppe for prosjektet: Elisabetta Vassenden(leder), Kristian Lønø, Terje Risberg og Lise Styrk Hansen.

Følgende referansegruppe for prosjektet er etablert: S.260 Arbeidsmarkedsstatistikk, Dag Rønningen, S.310 Demografi- og levekårsforskning, Trude Lappegård, S.320 Befolkning- og utdanning, Terje Risberg og Tor Jørgensen, S.330 Helsestatistikk, Jorun Ramm, S.340 Intervjuundersøkelser, Bengt Lagerstrøm, S.350 Levekårsstatistikk, Jorunn Lajord, S.370 Folke- og boligstelling, Harald Utne, S.420 Inntekts- og lønnsstatistikk, Harald Lunde, S.510 Offentlig økonomi og personmodeller, Dennis Fredriksen og S.720 Statistiske metoder og standarder, Anne Gro Hustoft.

Kirke- utdannings- og forskningsdepartementet har vist stor interesse for prosjektet. Både KUF og NIFU vil bli bedt om å gi konkrete innspill til prosjektet. KUF har antydnet at de kan være interessert i å støtte prosjektet økonomisk for å forsere tidsplanen.

Ressursrammer Hvilke økonomiske, tidsmessige og personalmessige ressursrammer foreligger.

Prosjektet kan deles i to deler når det gjelder fremdriftskrav og ressursrammer:

Del 1. Tilrettelegging av dataene i basen og system for uttak av datafiler som kan viderebearbeides til tabeller. Denne delen av prosjektet skal være klar 1. desember 2001. Forventet ressursbruk i del 1 er 2500 timeverk (21 månedsværk) på fagsiden(s320) og 4300 timeverk (36 månedsværk) på IT-siden, fordelt på 24 månedsværk på s303 og 12 månedsværk på s320. Gruppe for metoder bidrar i tillegg med 300 timeverk.

Del 2. Etablere et system for uttak av tabeller. Denne delen av prosjektet skal være ferdig 1. desember 2002 og forventet ressursbruk er 600 timeverk (5 månedsværk) på fagsiden og 12 månedsværk på IT-siden, fordelt med 6 månedsværk på s320 og 6 på s303.

Planlagt første publisering fra databasen vinteren 2001/2002.

Prosjektets prioritet ...også i forhold til andre oppgaver.

Prosjektet har sammen med ferdigstilling av den løpende utdanningsstatistikken høyeste prioritet på gruppe for universitets- og høgskolestatistikk på seksjonen.

PROSJEKTSKRIV fase 2

	Adm. dir.		
Fra:	Seksjonsleder Elisabetta Vassenden	Seksjon 320	Dato 21.09.01

Prosjektnavn Nasjonal utdanningsdatabase, Fase 2	Produktnr. 0330
Oppdragsgiver SSB	
Ansvarlig (utførende) enhet Seksjon for befolknings- og utdanningsstatistikk	
Prosjektleder Torill Vangen	

Se Statistisk sentralbyrås håndbøker 58 «Håndbok i prosjektstyring»

Bakgrunn

Her beskrives kort hvorfor/hvordan prosjektet har oppstått.

Prosjektet er del 2 av Nasjonal utdanningsdatabase (Nudb). Del 1 av Nudb skal være ferdig 1.12.01. Denne første delen består i å utvikle en prøvebase - en pilot der det skal tas ut en del data og tabeller til bruk for testing mot utdanningsinstitusjoner og tidligere publisert utdanningsstatistikk. Aktuelle instanser som KUF og NIFU vil få en aktiv rolle i arbeidet med å kvalitetssikre dataene, og å vurdere tallene i forhold til aktuelle problemstillinger, slik at det endelige produktet blir best mulig, både med hensyn til nytte og kvalitet.

Dette andre delprosjektet, som er fase 2 av databaseprosjektet bygger på arbeidet som er gjort i Fase 1. Arbeidet i Fase 2 vil i hovedsak bestå i:

- oppretting av ev. svakheter ved basen man har funnet ved testing av data, ev. forbedring av løsninger som ble valgt i del 1,
- utredning av muligheten for å inkludere data om studenter i utlandet fremover og bakover i tid,
- utvikling av et tilpasset uttakssystem for filer og tabeller fra basen, samt
- dokumentasjon av den nasjonale utdanningsdatabasen.

Hovedformålet med utdanningsbasen er å sammenstille utdanningsdataene for få et statistikkssystem som i større grad enn tidligere, enkelt gir muligheter til å gjennomføre bredere og større analyser av all den informasjonen utdanningsstatistikkfilene inneholder, spesielt gjennomstrømningsanalyser. Men en slik base med så stor detaljeringsgrad og store mengder data vil bli veldig omfattende og kompleks, og det vil være komplisert å få meningsfylte data ut av basen. Det må derfor utvikles et uttakssystem som tilrettelegger for uttak av tabeller og filer fra basen på en relativt enkel måte.

Uttak fra basen skal kunne skje på følgende måter:

- Uttak av enkle tall og spørringer kan utføres av alle som har litt kjennskap til SQL
- Uttak av mer kompliserte spørringer, eller spesialiserte oppdrag utføres av såkalte 'superbrukere', personer som har spesialisert seg på uttak fra basen via SQL
- Uttak av ofte forekommende spørringer og tabeller kan skje via et relativt enkelt uttakssystem som er tilrettelagt for ett sett med standardiserte spørringer. Her kan brukeren gå inn og krysse av for hva slags data han/hun ønsker.

Formål Forslag til målsettinger for prosjektarbeidet.

Formålet med dette delprosjektet er altså å lage et hensiktsmessig uttakssystem til Nudb. Et slikt uttakssystem vil gjøre det enkelt for fagpersoner å ta ut data/tabeller eller filer, f.eks. til bruk for publisering, eller data som det er jevnlig etterspørsel etter.

Utrede mulighetene for å få inn opplysninger om studenter i utlandet, både fremover og bakover i basen, for å kunne tette 'huller' i utdanningsforløp som oppstår når studenter har et studieopphold i utlandet, og dermed kunne komplettere dataene i basen.

-
- Fase 2 omfatter også oppretting av ev. svakheter ved basen man har funnet ved testing av data, og ev. forbedring av løsninger som ble valgt i del 1. Dette for å få en best mulig løsning til slutt.
-
- En viktig del av fase 2 er grundig dokumentasjon av basen. Med en så sammensatt og detaljert base er det viktig med god dokumentasjon for å kunne få full nytte av den. Basen skal dokumenteres i metadatabasen som er utviklet ved FD-trygd.

Faglige problemstillinger Vektlegg kort faglige forutsetninger og problemstillinger. Dette punktet utformes i detalj senere, i forbindelse med utformingen av milepælplanen.

Utdanningsdatabasen skal være en forløpsdatabase i Oracle. En tilsvarende modell er allerede i bruk ved FD-Trygd-basen i SSB. I motsetning til utdanningsbasen omfatter ikke FD-Trygd-basen tabelluttaksverktøy. Det skal utvikles et brukervennlig og fleksibelt tabelluttaksverktøy som er tilpasset de mest alminnelige tabellene til årlig publisering og ofte forekommende etterspørsler. Dette skal kunne brukes av fagpersoner som har begrenset kjennskap til SQL, slik at de raskt og forholdsvis enkelt skal kunne få ut data de har bruk for.

Muligheten for å ta inn data om utdanning som foregår i utlandet er etterspurt, for å kunne dekke noen av 'hullene' i forløpene som forekommer i perioder når studentene er utenlands. Det skulle være mulig å få disse dataene fra Lånekassa, men tidligere har dataene kommet for sent til å komme inn i årgangsfilene over igangværende utdanning. Det skal også ses på om det er mulig å legge inn data for dette bakover i tid.

Ved testing av data må det sjekkes om dataene er konsistente med tidligere publisert statistikk. I tillegg bør man se på om de mest hensiktsmessige løsningene valgt i Fase 1 i forhold til hva man vil ha ut av basen. Eventuelle justeringer i systemet gjøres i denne fasen.

Det er viktig å få dokumentert alle sider ved utdanningsbasen. Basen skal dokumenteres i metadatabasen som er utviklet ved FD-trygd. Det er produsert en del dokumenter underveis i arbeidet, der ulike alternativer av valg som er gjort er utredet. Det bør også skrives et dokumentasjonsnotat i papirformat.

Interessenter og deltakere Hvem er oppdragsgiver, og hvem er prosjektleder. Hvem er øvrige interessenter i eller utenfor Statistisk sentralbyrå. Hvilke fagområder eller organisatoriske enheter må bidra.

Prosjektet er i utgangspunktet et omleggingsprosjekt internt på seksjon for befolknings- og utdanningsstatistikk, men de fleste seksjoner i SSB som arbeider med individbasert statistikk vil være interessenter i prosjektet. Statistikkgrunnlagene fra utdanningsstatistikken benyttes i dag blant annet av s260, s310, s330, s340, s350, s420 og s520. Det vil også finnes eksterne interessenter i

form av forskere og andre som får utdrag av dagens filer for analyse og forskningsoppgaver.

Prosjektgruppen for Fase 2 består av Torill Vangen (prosjektleder), Lise Styrk Hansen, Astri Høiskar, Kjetil Digre, Arild Rognan m/flere fra fagsiden. IT- siden består av Anders Akselsen og Knut Inge Bøe m/flere fra 303 samt Anne Marie Holseter fra 320.

Følgende utgjør styringsgruppe for prosjektet: Elisabetta Vassenden(leder), Mari Sandelien, Kjetil Digre, Terje Risberg, Terje Krokstad og Lise Styrk Hansen.

Følgende referansegruppe for prosjektet er etablert: S.260 Arbeidsmarkedsstatistikk, Tor Jørgensen, S.310 Demografi- og levekårsforskning, Trude Lappegård, S.320 Befolkning- og utdanning, Terje Risberg, S.330 Helsestatistikk, Jorun Ramm, S.340 Intervjuundersøkelser, Bengt Lagerstrøm, S.350 Levekårsstatistikk, Øyvind Sivertstøl og Sandra Lien, S.370 Folke- og bolig telling, Harald Utne, S.420 Inntekts- og lønnsstatistikk, Harald Lunde, S.510 Offentlig økonomi og personmodeller, Dag Rønningen og S.720 Statistiske metoder og standarder, Anne Gro Hustoft.

Den eksterne referansegruppa består av Kirke- utdannings- og forskningsdepartementet/Læringssenteret, NIFU og Frisch senteret. Disse vil i tillegg bli bedt om å gi konkrete innspill til prosjektet.

Ressursrammer Hvilke økonomiske, tidsmessige og personmessige ressursrammer foreligger.

Fremdriftskrav og ressursrammer for Fase 2:

Testing av data, samt oppretting av svakheter og feil i basen. Forventet ressursbruk: 15 månedsverk på fagsiden og 13 på IT-siden, fordelt på 10 månedsverk på s303 og 3 på 320

Etablere et system for uttak av tabeller. Denne delen av prosjektet skal være ferdig 1. desember 2002 og forventet ressursbruk er 8 månedsverk på fagsiden og 18 månedsverk på IT-siden, fordelt med 9 månedsverk på s320 og 9 på s303.

Utredning av mulighetene for å få inn opplysninger om studenter i utlandet, både fremover og bakover i basen, samt ev. implementering av dette: Forventet ressursbruk: 5 månedsverk på fagsiden (s320) og 4 månedsverk på IT-siden fordelt på 3 månedsverk på s303 og 1 månedsverk på 320. Det er interesse fra bl.a. Frish senteret for å få denne delen av prosjektet utført, og dette vil forsøkes finansiert eksternt.

Dokumentasjon: 3 månedsverk på IT-siden, s303, 12 månedsverk på fagsiden

Totalt: 6,5 årsverk, fordelt på 3,4 årsverk på fagsiden og 3,1 årsverk på IT-siden (2 årsverk på s303 og 1,1 på 320).

Prosjektets prioritet ...også i forhold til andre oppgaver.

Prosjektet har høy prioritet på gruppe for universitets- og høgskolestatistikk på seksjon 320.

Begrepsforklaringer

NUDB er et datavarehus som inneholder forløpsdata utviklet etter "FD-Trygd-modellen". Det har i løpet av de årene FD-Trygd har vært under utvikling vært et stort behov for å sikre at fag- og IT-siden i prosjektet har kommunisert vha. like og entydige begreper. Mens noen av begrepene som er utviklet har klare paralleller til mer tradisjonelle datavarehus, er andre begreper helt spesifikke for den type data vi har i SSB og den modellen som vi har utviklet. Uten et slikt felles begrepsapparat ville oppgaven med å bygge opp NUDB vært tilnærmet umulig. Nedenfor er en alfabetisk ordnet oversikt over de ord og uttrykk som er benyttet under arbeidet med å danne forløp i NUDB ut fra grunnlagsregistre lagret som tradisjonelle årsfiler.

Leseveiledning:

Noen av begrepene vil være kjent fra andre sammenhenger i SSB og ellers. De fleste begrepene er felles for hele "FD-Trygd-modellen", mens noen få er spesifikke for selve NUDB-prosjektet. De som er spesifikke for NUDB er merket med: (NUDB). Det vil i tillegg forekomme begrepsforklaringer som benytter begreper som er forklart andre steder i denne listen. Disse begrepene vil være skrevet i *kursiv* (Eksempel for begrepet Avgang: "En avgang er en av tre type *hendelser* som registreres...").

Aktiv bestand

En aktiv bestand er alle de personer i en populasjon som på et gitt tidspunkt er i gang med noe (er i arbeid, under utdanning, mottar sosialhjelp, mottar arbeidsledighetsstrygd etc.). I "FD-Trygd-modellen" vil vi si at dette er tilfeller som er løpende på et gitt tidspunkt. Når vi transformerer årgangsfiler til forløpsdata vil den aktive bestand bestå av alle de tilfeller som til enhver tid er representert med en record i hovedtabellen for det aktuelle statistikkområdet.

Avgang

En avgang er en av tre typer hendelser som registreres med egne records i NUDB (de to andre hendelsene er tilgang og endring). En avgangsrecord inneholder datoer for når tilfellet blir avsluttet. I tillegg inneholder recorden selvfølgelig alle andre variabler (internvariabler etc.) som vi ønsker å benytte til statistikkproduksjon, analyse etc. senere. En avgangsrecord vil alltid være lik den foregående recorden for tilfellet mhp. de daterte variablene. Avganger dateres vha. REGDATO og AVGDATO (og en gruppeavgangsdato for de temaområdene der vi opererer med hovedgrupper).

For NUDB: Opplysninger om når et tilfelle avsluttes får vi fra avsluttafilene. For de tilfellene som vi mangler avslutningsopplysninger for, avslutter vi selv med en egen kode (UTFALL = 1) for at ingen tilfeller skal få løpe evig. Kun REGDATO og AVGDATO vil endre seg i tillegg til variabler som er avgangsspesifikke (UTFALL etc). Avgangen dateres vha. variabelen STOPPD på grunnlagsfilene.

Avgangstabell

Dette er en teknisk hjelpetabell på databasen som inneholder de siste avgangssopplysningene for hvert tilfelle som er registrert i datavarehuset. Tabellen benyttes bl.a. for konsistenssjekking og datokontroll ved danning forløpsdata når nye årganger kommer til. I tillegg kan den benyttes for å kunne kontrollere i ettertid at innholdet i datavarehuset er som planlagt. Tabellen benyttes ikke ved uttak av data fra datavarehuset og har ingen betydning uttakssystemet overhodet.

Avgdato

Dette er en teknisk hjelpevariabel. Daterer når et tilfelle avsluttes. Settes når vi får en avgangssopplysning fra datamaterialet, eller når vi avbryter et tilfelle som ikke avsluttes på ordinær måte i datamaterialet. AVGDATO vil ligge blank ellers (NULL) på alle andre records enn for avgangsrecorden for det samme tilfellet. Det lages en helt ny record for hver gang vi registrerer en avgangssopplysning, og hver avgangsrecord vil alltid være lik den foregående recorden for tilfellet mhp. de daterte variablene. Kun AVGDATO og REGDATO vil endre seg i tillegg til de variabler som er avgangsspesifikke.

For NUDB: Daterer når et kurs- eller utdanningstilfelle avsluttes.

Datavarehus

Et datavarehus (eller en forløpsdatabase) er "en logisk sammensetting av data som er hentet fra allerede eksisterende operasjonelle systemer (les: grunnlagsfiler for offisiell statistikk), der data i forløpsdatabasen er strukturert og tilrettelagt for analyse av sammenhenger mellom fortiden, nåtiden og fremtiden". En viktig egenskap ved datavarehus er at alle tidligere situasjoner som har eksistert på grunnlagsfilene vil kunne gjenskapes i ettertid. For å få til dette er det et ufravikelig krav at data som legges inn i et datavarehus ikke skal revideres når de er blitt en del av datavarehuset.

Datering

Alle hendelser som "inntreffer" når vi sammenlikner data fra ett årssett, med data fra et foregående årssett, blir tatt vare på. Under oppbygging av forløpsdata er det kun opplysninger som endrer det vi allerede vet i forløpsdatabasen fra før av som det er interessant å registrere. Dette gjøres ved å datere nye hendelser på nye records. Dersom flere hendelser inntreffer på samme tid (at flere daterte variabler endrer verdi på samme tid), så registreres disse på en og samme record.

For NUDB: Ved hendelsen *tilgang* settes f.eks. TILGDATO, REGDATO pluss at alle daterte variabler dateres vha. STARTD. Ved endring på en eller flere av disse daterte variablene settes REGDATO og de endrede variablene etter gitte regler, de øvrige dateringene endrer ikke verdi. Ved avgang settes AVGDATO og REGDATO lik STOPPD.

Datert variabel

Et datert variabel er en variabel fra miniregisteret eller en variabel som er utledet fra en eller flere andre variabler på miniregisteret, hvor endringer på variabelen er datert til når endringen fant sted. Dateringen av slike variable foregår på egne dateringsvariable.

Endring

Med endring menes her endring på en eller flere av de daterte variablene når data fra ett årssett sammenliknes med det foregående. Dersom det detekteres endringer på en eller flere av disse variablene genereres det en egen (endrings-)record som daterer når disse endringene fant sted. Senere, når neste årssett med data så sammelikes mot denne record, og nye endringer detekteres, vil det bli generert ennnå en endringsrecord som daterer disse endringene igjen. Det er kun daterte variable som har endret verdi som vil dateres på nytt, alle andre dateringer vil ligge urørt.

Forløpstabell

En forløpstabell er en Oracle-tabell som inneholder data tilrettelagt for uttak av forløpsanalyse, tverrsnittsanalyse etc.

For NUDB: I NUDB har vi to slike tabeller knyttet til utdanningsdata; en for forløp på kursnivå (F_UTD_KURS) og en for forløp på personnivå (F_UTD_PERSON). Hver person som har tatt en eller annen utdanning i Norge vil ha en eller flere records i disse tabellene. De nevnte recordene vil angi når en persons utdanningsaktivitet startet, når den evt. ble endret og når den evt. ble avsluttet. Hver hendelse som inntreffer vil registreres med nye records i disse tabellene. For hver årgang (en årgang er en igangvfil pluss en avsluttafil pluss en BHU-fil for et år) genereres ca. 1-2 millioner nye records ut i tabellen F_UT_KURS, og ca. 1-1,5 millioner nye records ut i tabellen F_UTD_PERSON.

Grunnlagsregister

Grunnlagsregistre er de registre som har vært benyttet som bakgrunnsfiler for produksjon av statistikk.

For NUDB: Alle igangværende-, avslutta-, og BHU-filer er grunnlagsregistre. Disse filene er benyttet som grunnlag for å danne sk. miniregistre. Miniregistrene er så benyttet som input til programmene som har dannet forløp i NUDB.

Gruppeavgang

En gruppeavgang er en avgang fra en hovedgruppe; f.eks. videregående utdanning, eller høyere utdanning. Gruppeavganger etterfølges ofte av gruppetilganger, dvs. tilgang til en ny hovedgruppe. Gruppeavganger genererer alltid en record i datavarehuset som markerer hendelsen. I tillegg vil en egen hendelseskode markere at recorden er en avgangsrecord (med kodeverdi = 0). En gruppeavgang pluss en gruppetilgang kalles for en overgang.

For NUDB: Gjelder for tilfeller på personnivå (utdanningstilfeller). Variablene UTDAVG og REGDATO vil datere hendelsen på disse recordene, og andre siffer i UTDKODE vil markere at dette er en avgang (kodeverdi = 0). Gruppeavganger dateres vha. STOPPD

Gruppetilgang

En gruppetilgang er en tilgang til en hovedgruppe; f.eks. høyere utdanning. Gruppetilganger etterfølger alltid gruppeavganger. Gruppetilganger genererer alltid en record i datavarehuset som markerer hendelsen. I tillegg vil en egen hendelseskode markere at recorden er en tilgangsrecord (med kodeverdi = 1). En gruppeavgang pluss en gruppetilgang kalles for en overgang.

For NUDB: Gjelder for tilfeller på personnivå (utdanningstilfeller). Variablene UTDTILG og REGDATO vil datere hendelsen på disse recordene, og andre siffer i UTDKODE vil markere at dette er en tilgang (kodeverdi = 1). Gruppetilganger dateres vha. STARTD.

Hendelse

I NUDB (og andre forløpssystemer basert på FD-Trygd-modellen) vil det være registrert tre typer hendelser i datavarehuset. De tre er: tilgang til et tilfelle, avgang fra et tilfelle og endring i daterte variablene for et tilfelle. Hver hendelse registreres i NUDB med en egen record som daterer hendelsen.

Hjelpetabell

Hjelpetabeller er Oracle-tabeller som brukes ved danning av forløp i datavarehuset. Dette er tekniske tabeller for programmene som danner forløp og som ikke skal benyttes til statistikk/analyse.

Hovedtabell

En Oracle-hjelpetabell som til enhver tid inneholder et øyeblikksbilde av de tilfellene som er løpende (ikke avsluttet). Alle nye data fra igangværende-, avslutta-, og BHU-filene kobles mot denne for å finne ut om disse nye dataene representerer en tilgang til et nytt tilfelle, en avgang for et gammelt tilfelle, eller om det ev. har skjedd endringer på noen av de daterte variablene for løpende tilfeller i forløpstabellene.

Inaktiv bestand

Benyttes om tilfeller som er avsluttet (alle records i avgangstabell).

For NUDB: Dette er kurs og utdanningstilfeller som ikke løper lenger, men som er avsluttet i datavarehuset.

Kurs (NUDB)

Et kurs er en enkeltutdanning med en bestemt NUS-verdi. Til et kurs hører alle hendelser (tilgang, endringer og avgang) som har inntruffet for en persons enkeltutdanning. Det er kurset, eller "enkelutdanningen" som er enheten, og en og samme person kan være i gang med flere kurs samtidig. Fra 1999 er avlagte studiepoeng registrert som et kurs. Disse har rectype 3.

Lastetabell

Dette er en temporær hjelpetabell som benyttes som mellomlager/mottakslager for de flate filene (med grunnlagsdata) som benyttes når forløpet bygges opp i datavarehuset.

For NUDB: Årsfiler for BHU-, igangv.- og avsluttafilene lastes fra flate filer og inn i laste-tabeller (også kalt load-tabeller).

Løpende tilfeller

At et tilfelle er løpende vil si at vi per dags dato ikke har registrert noen avslutning på det. Alle tilfeller som er løpende er registrert i en sk. hovedtabell i tillegg til i forløpstabellen. Løpende tilfeller i hovedtabellen benyttes bl.a. for å fastslå hva data fra nye årssett (For NUDB: BHU-, igangv.- og avsluttafilene) gir av nye opplysninger. Mao. om nye data viser at endringer har skjedd på en eller flere av variablene som vi daterer, eller om opplysningene avslutter et løpende tilfelle.

Match

Å matche to tabeller (eller filer) betyr ganske enkelt å koble en tabell mot en annen vha. en felles nøkkel (f.eks. FNR). Dersom en rad i en tabell matcher en (eller flere) rader i en annen tabell vha. en koblingsnøkkel, heter det at vi har en match (på fødselsnummer i dette tilfellet).

Miniregister

Miniregistre benyttes om de grunnlagsregistrene vi har bearbeidet for å kunne brukes som input til de programmene på databasesiden som danner forløp i datavarehuset. Miniregistrene er sjekket for dubletter, det er foretatt kontroll av fødselsnummer og datokonsistens er ivaretatt. I tillegg er variable sjekket over tid for å se at de inneholder samme type informasjon over hele perioden.

Nøkkel

Den eller de variablene som til sammen sikrer at vi har unike records i datavarehuset.

Overgang

Når en person skifter hovedgruppe (f.eks. fra videregående til en høyere utdanning) kaller vi dette for en overgang og genererer to records for å markere dette i NUDB; en avgangsrecord for den gamle hovedgruppen, og en tilgangsrecord for den nye hovedgruppen.

For NUDB: Gjelder for tilfeller på personnivå (utdanningstilfeller).

Regdato

Dette er en teknisk hjelpevariabel. Regdato fungerer som den siste oppdaterings-/endringsdato for recorden sett under ett, uavhengig om det er en tilgang, avgang eller endring, og angir således hvilken dato det sist skjedde noe på recorden. Benyttes mye ved uttak fra datavarehuset, men også når forløp skal dannes under oppbygningen av datavarehuset.

Record

En record er databaseuttrykket for en rad/linje med data. En record kan ha en eller flere kolonner (felt/kjennemerker/variabler).

Sumtabell

Tabellen inneholder variabler som ikke endrer seg over tid og som på grunn av bl.a. ønske om å spare plass og senke responstiden ved spørringer mot datavarehuset, er lagt i en egen tabell.

For NUDB: TAB_UTD_PERSON er en slik tabell. Tabellen inneholder bl.a. opplysninger om kjønn, fødselsår, fødeland, årstall for første fullføring av grunnskole, videregående skole, høyere utdanning osv.

Tidsbuffer

Beskriver det tidsrommet vi må vente med å levere ut data fra datavarehuset pga. at vi ikke vet om et tilfelle som løper skal avsluttes eller ikke, før neste års data er registrert.

Tilfelle

Et tilfelle beskriver alle hendelser for en person, (eller et kurs for NUDB), fra en aktivitet starter til denne aktiviteten er avsluttet.

For NUDB: her snakker vi om to typer av tilfeller; kurstillfeller og utdanningstillfeller. Et kurstillfelle er alle records som beskriver hendelsene for et kurs som en person tar. Et utdanningstillfelle inneholder alle records som beskriver hendelser omkring en persons utdanning, fra start på videregående skole til ev. avslutning på ett eller annet høyere nivå.

Tilgang

En tilgang er en av tre typer hendelser som registreres med egne records i NUDB (de to andre hendelsene er endring og avgang). En tilgangsrecord inneholder datoer for når tilfellet ble startet opp, i tillegg til alle andre variabler som vi ønsker å benytte til statistikkproduksjon, analyse etc. senere. Tilganger dateres vha. REGDATO, TILGDATO (og en gruppetilgangsdato for de temaområdene der vi opererer med hovedgrupper).

For NUDB: Tilgangen dateres vha. variabelen STARTD på grunnlagsfilene.

Tilgdato

Dette er en teknisk hjelpevariabel. Tilgangsdatoen settes når vi starter opp et nytt tilfelle. Datoen følger alle records for et tilfelle og endres aldri.

For NUDB: TILGDATO settes lik STARTD fra den record i årsfilen som gir oss opplysningen (litt forenklet).

Tverrsnitt

Over tid kan (vil) innholdet i variabler for en populasjon endre seg. Et tverrsnitt er et "frost" bilde av innholdet i disse variablene til denne populasjonen på et bestemt tidspunkt.

Variabeltyper

Hver enkelt variabel (kjennemerke) er kategorisert for lettere å avgjøre hvordan de skal behandles ved oppbygging av datavarehuset, ved uttak og ved utlevering av variabellister fra metadatabasen (MetadB). En internvariabel er en variabel fra grunnlagsregisteret eller en variabel som er utledet fra en eller flere variabler fra grunnlagsregisteret. Variabler som er knyttet til modellen er:

Variabeltyper	Kommentar	Eksempler
Id for statistisk enhet (fnr)	Kun FNR (fødselsnummer)	f_utd_kurs.FNR
Gruppe og/eller hendelse	Hovedgrupper og/eller hendelser	f_utd_kurs.HOVED, f_utd_kurs.KODE
Hendelsesdato (fom)	Kun REGDATO	f_utd_kurs.REGDATO
Hendelsesdato (tom)	Skal ikke benyttes (det finnes ingen slike variabler foreløpig)	Ingen
Tilfelledato	Variabler som daterer tilgang og avgang på tilfellenivå	f_utd_kurs.TILGDATO, f_utd_kurs.AVGDATO
Gruppetilfelledato	Variabler som daterer tilganger og avganger på (hoved-) gruppenivå	f_utd_kurs.UTDTILG, f_utd_kurs.UTDAVG
Kurstilfelledato	Benyttes kun av kurs-tabellen (F_UTD_KURS)	f_utd_kurs.TILGDATO, f_utd_kurs.KURSAVG
Internvariabel (datert)	(Intern-) variabler som endrer seg over tid og hvor hendelsene dateres i egne variabler (internvariabeldatoer)	f_utd_kurs.KLTRINN f_utd_kurs.HELDEL f_utd_kurs.KOMMNR
Internvariabel (ikke datert)	(Intern-) variabler som <u>ikke</u> endrer seg over tid og av den grunn ikke har noen hendelser som kan dateres	Settes en gang, endres ikke for tilfellet f_utd_kurs.NUS2000 f_utd_kurs.UTD
Internvariabeldato	Variabler som daterer hendelser for internvariabler	f_utd_kurs.KLTRINNDATO f_utd_kurs.HELDELDATO
Internvariabel (dato, datert)	Internvariabler som i seg selv er en dato, og som i tillegg dateres vha. en egen Internvariabeldato.	
Internvariabel (dato, ikke datert)	Internvariabler som i seg selv er en dato, men som <u>ikke</u> dateres vha. en egen Internvariabeldato.	Settes en gang, endres ikke for tilfellet
Merkingsvariabel	Variabler som inneholder kode for hvilke revideringer som evt. er foretatt på andre variabler	f_utd_kurs.MRK_FNR, f_utd_kurs.MRK_DL
Teknisk variabel	Ulike tekniske hjelpevariabler	f_utd_kurs.STATUS1, f_utd_kurs.STATUS5

Venstresensur

Venstresensur benyttes om tilfeller som er startet opp før "tidsregningen" for datavarehuset starter opp (For FD-Trygd før 01.01.1992). Venstresensur angis med en egen hendelseskode (kodeverdi = 3). For venstresensurerte tilganger vil det ikke være samsvar mellom de tekniske dateringsvariablene TILGDATO og REGDATO. For TILGDATO benyttes de "riktige datoene" for tilfellet (de som ligger på grunnlagsfilene), mens REGDATO vil settes til venstresensurdatoen. Dette gjøres for at det ikke skal bli forvirring om hvor langt tilbake i tid vi kan levere ut tverrsnitt, forløpsdata etc. fra datavarehuset.

For NUDB: Venstresensur benyttes om tilfeller som er startet opp før 197011.

Årgang (NUDB)

Variabelen AARGANG på miniregistrene angir årgangen for filene. BHU-filen per 1.10.1980 og igangværendefilen per 1.10.1980 vil ha AARGANG = 1980, mens avsluttafilen for perioden 1.10.1980 til 30.9.1981 vil ha AARGANG = 1981.

En statistikkfaglig utredning av dubletter på fnr/snr i Nudb

I diskusjonen om man skal bruke fødselsnummer (fnr) eller S-nummer (snr) som identifikasjonsnøkkel i basen ble det ytret ønske om å ta høyde for å rette opp dubletter i basen. Dette fordi man skal ha muligheten til å lage riktig forløp på flest mulig personer. Denne problemstillingen er uavhengig av om fnr eller snr velges som identifikasjonsnøkkel i basen.

Dubletter er i denne sammenhengen når samme person har hatt to ulike Fnr og to ulike Snr. F.eks. kan en person som har vært registrert bosatt i Norge i flere perioder, ha fått tildelt flere ulike fødselsnummer med tilhørende ulike Snr. Det vil dermed bli laget separate forløp i basen, som må linkes sammen.

Antall dubletter

I perioden 08.01.1999 til 29.03.2001 (ca. 27 måneder) var det 140 slike dubletter som har blitt oppdaget (årsakskode 58 i BEBAS). Dette utgjør gjennomsnittlig ca. 62 i året.

- 39 personer er under 16 år, og vil ikke ha kommet inn i utdanningsstatistikken enda, dvs. at problemet vil være løst før de kommer inn i basen.
- 101 personer er 16 år eller eldre. Det er disse som vil utgjøre et problem i individstatistikken (dette utgjør gjennomsnittlig ca. 44 personer per år).

Disse 101 er fordelt på følgende landbakgrunn:

Norske:	17
Andre nordiske land:	30
Andre europeiske land:	18
Asiatiske land:	24
Afrika/Amerika/ Australia:	12

Av disse 101 er det kun 43 som er registrert bosatt per 1.1.2000 med gammelt eller nytt fnr. Problemet synes i første rekke å være utenlandske personer som kommer til Norge på nytt etter å ha vært ute av landet en periode, ofte har de også endret navn, eller navn er registrert forskjellig.

Av de 101 er det

- 14 personer som har vært registrert i utdanning (på årgangsfilene), siden 1974 med sitt nye fødselsnummer, og
- 8 som har vært registrert i utdanning på sitt gamle fødselsnummer, siden 1974

Ingen av disse er registrert med både gammelt og nytt fnr.

For å kunne vurdere hvorvidt det vil være hensiktsmessig å ta høyde for å rette opp disse dublettene i basen, vil det være relevant å ta i betraktning hvor mange man ikke vil klare å lage forløp på totalt, dvs. hvor mange som ikke har gyldig fødselsnummer.

Antall med ugyldig fødselsnummer:

Når det gjelder kvaliteten på fødselsnummer i statistikken for øvrig er det et visst antall elever/studenter hvert år med ugyldig fødselsnummer, der man ikke klarer å lete opp riktig fødselsnummer gjennom maskinell eller manuell fødselsnummerleting. De siste 10 årene har antallet på elev/studentfila vært følgende:

Per 1.10.2004 var det til sammen 3958 uten gyldig fnr.

Per 1.10.99 var det til sammen 3513 uten gyldig fnr.

Per 1.10.98 var det til sammen 4332 uten gyldig fnr.

Per 1.10.97 var det til sammen 4000 uten gyldig fnr.

Per 1.10.96 var det til sammen 3408 uten gyldig fnr.
Per 1.10.95 var det til sammen 3172 uten gyldig fnr.
Per 1.10.94 var det til sammen 2973 uten gyldig fnr.
Per 1.10.93 var det til sammen 2667 uten gyldig fnr.
Per 1.10.92 var det til sammen 2468 uten gyldig fnr.
Per 1.10.91 var det til sammen 2705 uten gyldig fnr.
Per 1.10.90 var det til sammen 2473 uten gyldig fnr.

Når det gjelder fordelingen av disse på elever/studenter, og nordmenn/utlendinger kan denne oversikten fra 1988 gi en pekepinn:

- Studenter per 1.10.88 uten fnr. - 614 med 'utenlandsk-klingende' navn og 426 med 'norsk-klingende' navn
- Elever per 1.10.88 uten fnr - 217 med 'utenlandsk-klingende' navn og 344 med 'norsk-klingende' navn

Årsakene til det høye antallet med ugyldig fødselsnummer er flere:

- Utdanningsinstitusjonene leverer data uten fødselsnummer
- Utdanningsinstitusjonene leverer data med feil i fødselsnummer
- Utenlandske elever/studenter, som bare er på utveksling et eller to semestre, vil ofte ikke få tildelt noe fødselsnummer.
- Utdanningsinstitusjonene har tidligere laget fødselsnummer som ser gyldige ut, for å få lagt studentene inn i sine studentadministrative systemer.

Det bør gjøres en innsats mot utdanningsinstitusjonene for å få disse til i størst mulig grad å levere data med gyldige fødselsnummer. Institusjonene bør oppfordres til i minst mulig grad å legge inn studenter uten gyldig fnr, kreve at studentene oppgir gyldig fnr, og sørge for at utenlandske studenter i størst mulig grad skaffer seg gyldig fødselsnummer til de skal registreres.

Det er utarbeidet en rutine for automatisk fødselsnummerleting i bearbeidingsystemet for utdanningsstatistikken, denne ser ut til å være en forbedring fra den tidligere manuelle fødselsnummerletingen. Antallet med ugyldige fødselsnummer har gått noe ned, men en kan ikke regne med at tallet på ugyldige fødselsnummer kommer til å synke nevneverdig de kommende årene.

Konklusjon

For å kunne vurdere hvorvidt det vil være hensiktsmessig å ta høyde for å rette opp disse dublettene i basen ut fra et faglig synspunkt, vil det være relevant å ta i betraktning hvilken gevinst det vil ha i form av forbedrede utdanningsopplysninger totalt sett.

Som et prinsipp bør det være et mål i basen å ha flest mulig korrekte fødselsnummer, for å kunne lage flest mulig korrekte forløp, og få best mulig kvalitet på statistikken. I lys av dette er det et problem at det finnes dubletter i basen.

Antallet dubletter på fnr/snr utgjør ca. 40 stykker i året over 16 år. Selv om dette totalt sett er en liten andel, vil disse akkumuleres fra år til år. Det er også grunn til å tro at stadig flere vil ta utdanning, med det økende fokus på livslang læring.

Det er stort tidspress i dette prosjektet, og med tanke på dette bør det utredes om det kan gjøres i en senere fase av prosjektet å rette opp disse dublettene. Når det gjelder hvor mye arbeid det vil medføre på den tekniske siden vil dette bli utredet av IT-siden i prosjektet.

En slik IT-faglig dublettløsning i en forløpsbase vil kunne gjenbrukes på andre fagområder senere. Det har bl.a. tidligere vært drøftet å lage en hendelsesbase/forløpsbase for befolkningsstatistikk, og da vil en slik rutine være nyttig.

En IT-faglig utredning av dubletter på fnr/snr i Nudb

IT-teknisk kan problemene som oppstår ifm. dubletter løses på tre ulike sett: ved hjelp av manuell-, semi-manuell-, eller maskinell oppretting. Det vil i denne forbindelse være naturlig å foreta en kost/nytte-analyse av de ulike løsningsforslagene før en avgjør hvilken metode en velger. De ulike løsningsforslagene som er beskrevet nedenfor er til slutt analysert i en kostnadsmatrise.

Manuell oppretting

Dette kan foregå ved hjelp av egnede databaseverktøy som SSB råder over i dag og som gjør det mulig å identifisere, hente fram og rette opp data i regnearklignende grensesnitt (Golden32 og Toad). I praksis vil en slik oppretting foregå ved at en ved hjelp av (papir-)lister med dubletterte Snr/Fnr klikker og peker seg fram i programvaren som nevnt ovenfor, til en sitter igjen med forløpsdataene for hvert enkelt individ som skal rettes opp. Ved å sortere dataene etter definerte nøkler kan en deretter visuelt sjekke recordene, og ved behov rette opp datoer og andre verdier i NUDB når disse er i konflikt med gjeldende konsistensregler for datamodellen.

Semi-manuell oppretting

Innebærer at en utvikler en applikasjon med et enkelt grensesnitt som gjør det mulig å se på /evaluere de tilfeller hvor en har dubletter. Applikasjonen kan ha noen enkle innebygde funksjoner som automatisk identifiserer noen av de mest vanlige konsistensproblemene som kan inntreffe når en opererer med dubletterte identifikasjonsnøkler. Selve opprettingen foregår ved at en manuelt retter/setter inn nye verdier i NUDB for de individene dette gjelder på en liknende måte som er beskrevet under manuell oppretting.

Maskinell oppretting

Her utvikles det en batch-rutine som identifiserer de dubletterte Snr/Fnr i NUDB vha koblinger mellom nummerene og deres data i NUDB. Deretter sammenlignes og analyseres data for hvert enkelt individ automatisk. Programmet avgjør deretter hvilke endringer/slettinger som skal foretas i hvert enkelt tilfelle. For å få til dette må programmet kunne holde oversikt over alle data for de dubletterte individene i NUDB samt alle reglene som er benyttet for å bygge opp forløpet for disse. For å sikre konsistens i NUDB (både mellom hver enkelt tabell og innad i hver tabell) kreves det at en på forhånd kan beskrive alle kombinasjoner av situasjoner som kan oppstå når data fra et individ med to eller flere forløp skal settes sammen til ett.

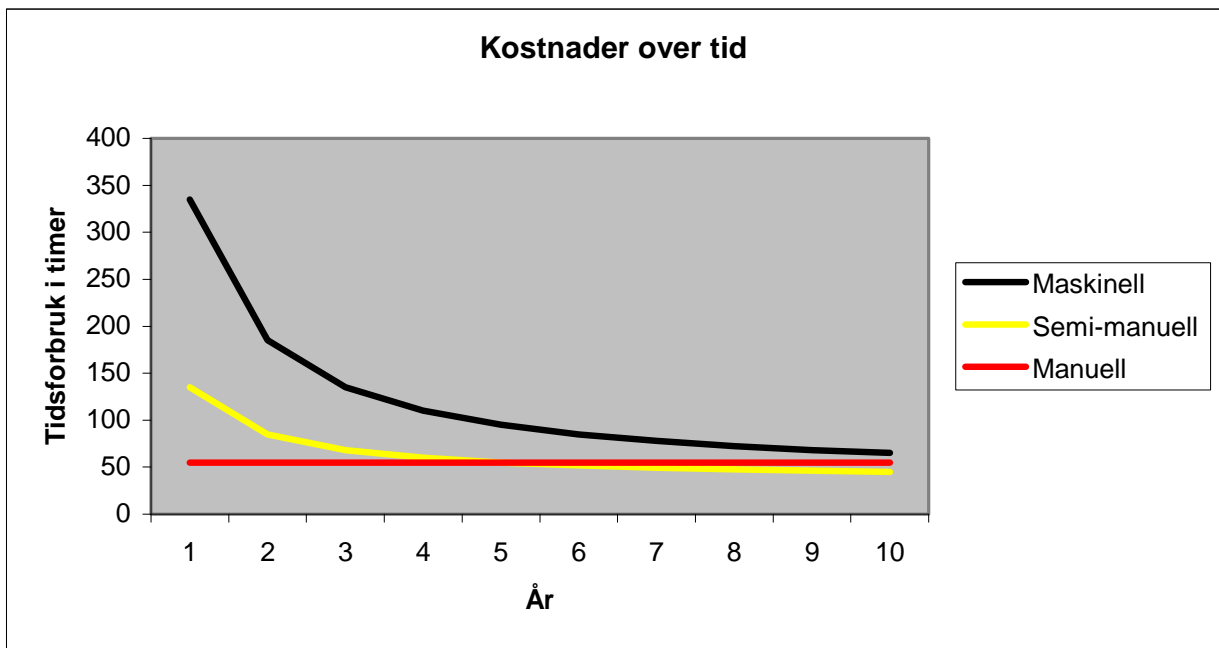
Kost/nytte-analyse

Nedenfor er en kostnadsanalyse som er gjort på fire sentrale områder. Selv om timeverkene er grove anslag er det rimelig å anta at de relative innbyrdes forhold som kommer fram er korrekte:

Kostnadsmatrise

Beskrivelse	Manuell	Semi-manuell	Maskinell	Begrunnelse
IT-sidens kostnader ifm. utvikling av en dublettløsning -engangskostnad	0t	100t	300t	NUDB inneholder svært mange ulike variabler, samt regler for å håndtere disse over tid. Alt henger sammen med alt i NUDB og en maskinell løsning vil måtte ta høyde for dette. I en semi-maskinell løsning vil en kombinasjon av manuell vurdering av alle involverte records kombinert med manuell oppretting av disse, spare tid.
IT-sidens vedlikeholds-kostnader -årlige kostnader	0t	15t	25t	Dersom datamodellen endres og/eller at NUDB utvides med flere variabler i årene som kommer, vil en måtte revidere programmer som arbeider mot NUDB. Erfaring tilsier at et komplisert databasescript krever mer å vedlikeholde enn en enklere Forms-applikasjon.
IT-sidens driftskostnader (opplæring/ bistand ifm. regelverk) -årlige kostnader	15t	10t	5t	I en situasjon der fagsiden foretar en manuell/semi-manuell oppretting vil det fortsatt være behov for IT-faglig veiledning ang. konsistensspørsmål i fm. datamodellen og det regelverk som er benyttet under oppbygningen av den.
Fag-sidens kostnader ifm. årlig retting av dubletter -årlige kostnader	40t	10t	5t	I en situasjon der IT-siden foretar en maskinell oppretting vil fagsiden fortsatt måtte bidra med faglig veiledning når skjønnsmessige avgjørelser skal tas. En semi-manuell oppretting vil kunne foretas raskere enn en manuell oppretting da førstnevnte vil ha en del innebygde rutiner som letter arbeidet noe.

Kostnader over tid



Konklusjon

NUDB bygges opp vha. til dels kompliserte rutiner som håndterer sammenhengen mellom allerede innlagte data og nye data som tilføres årlig. Det er en grunnleggende egenskap ved denne type løsninger at alt henger sammen med alt og at eldre data i databasen er låst for oppdateringer. Dette er en nødvendighet der en ønsker å se flere tiår med til dels ulike data i sammenheng samtidig som en ønsker å kunne foreta ad-hoc spørringer som gir svar på kompliserte spørsmål. Det er med andre ord av avgjørende betydning at en bygger opp databasen vha. ferdig reviderte data. I praksis viser det seg imidlertid at selv ferdig reviderte data må kunne revideres igjen på et senere tidspunkt. Dersom antall records som er involvert i en slik revisjon ikke er for stort, eller at graden av revisjon ikke er for komplisert, lar dette seg gjøre. Enkelte ganger kan det være hensiktsmessig å gjøre dette maskinelt, mens andre ganger vil en manuell oppretting gi både det beste faglige resultatet samtidig som det er raskere å utføre.

Når vi ser på problemene med dubletter i NUDB har vi en situasjon hvor det er et relativt lavt antall individer involvert, men et komplisert regelverk som håndterer disse. I et slike tilfelle kan det være vanskelig å lage fullgode maskinelle løsninger som ivaretar de faglige krav som er satt. Det vil etter vår mening ikke være mulig å garantere 100% riktige opprettinger av dubletter vha. den maskinelle metoden. Det vil være svært vanskelig å forutse alle mulige kombinasjoner som kan oppstå når en forsøker å sette sammen data for et individ som har vært betraktet som to under tillegg i NUDB, fordi alt henger sammen med alt. Dublettene kan f.eks. ha forårsaket tidsmessige overlapp hvor datoer og andre grunnleggende opplysninger i NUDB er feil. Men også uten tidsmessig overlapp er det sannsynlig at grunnleggende opplysninger er i konflikt med hverandre slik at skjønsmessige valg må foretas.

Det lave antall dubletter per år vil etter vår mening ikke kunne forsvare det å legge ned ressurser i å utvikle (og vedlikeholde) en applikasjon/databasescript som beskrevet i notatet. IT-kontoret anbefaler derfor at en velger en manuell løsning i første omgang og gjør seg noe erfaring med denne metoden før en ev. vurderer en semi-manuell løsning.