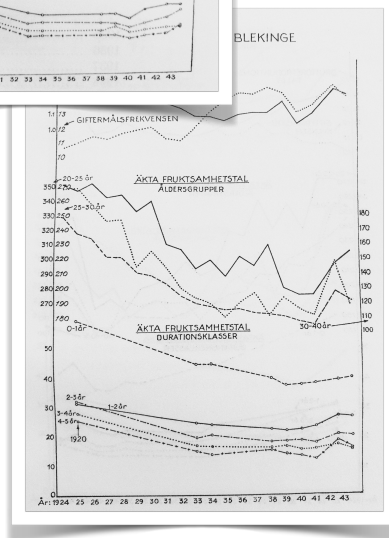
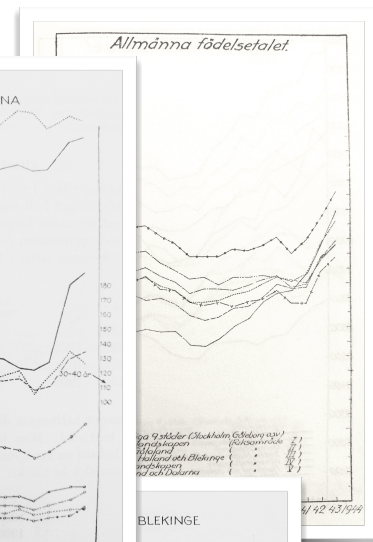
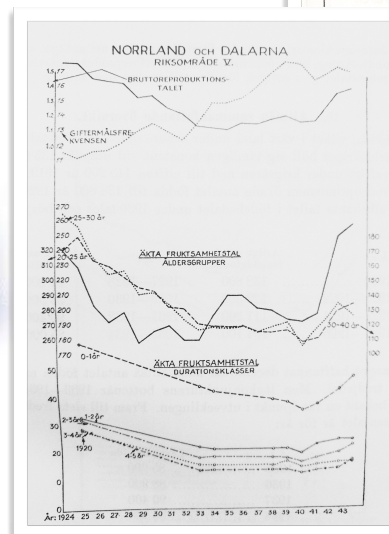




SwedPop finansieras av Vetenskapsrådet

2018-2022 och stöds av ett konsortium

bestående av följande parter:

- **Umeå universitet: Demografiska Databasen (DDB)** Enheten för demografi och åldrandeforskning CEDAR. Livsförloppsdata för befolkningen i ca 100 församlingar i olika svenska regioner 1650-1960. Folkräkningen 1950. Koordinator för projektet.
- **Göteborgs universitet: Göteborgs Befolkningspanel (GOPP)** Institutionen för ekonomi och samhälle. Fyraårspaneler för ett slumpvis urval individer bosatta i Göteborg under något av panelåren 1915-1943.
- **Lunds universitet: Skånes Ekonomisk-Demografiska Databas (SEDD)** Centrum för ekonomisk demografi. Livsförloppsdata för befolkningen i utvalda skånska församlingar 1680-1968.
- **Riksarkivet: SweCens-databasen** Individdata från svenska folkräkningar 1880-1950. Tvärsnitt vart tionde år för hela riket. Sveriges Dödbok (via avtal med Sveriges Släktforskarförbund).
- **Stockholms stadsarkiv: Rotemansarkivet** Livsförloppsdata för Stockholms befolkning 1877-1926.

SwedPop: Svenska befolkningsdatabaser för forskning 2018-2022

1. Inledning

Under 2017 beviljades den nationella infrastrukturen SwedPop bidrag från Vetenskapsrådet inom ramen för utlysningen *Bidrag till infrastruktur av nationellt intresse*. Projektets syfte är att utveckla och samordna de viktigaste historiska befolkningsdatabaserna i Sverige till en gemensam resurs, som väsentligt kommer att öka, förbättra och förenkla forskares tillgång till högkvalitativa svenska befolkningsdata för forskning. Databaserna samordnas och tillgängliggörs i en ny infrastruktur, SwedPop, som bland annat kan användas för jämförande studier av hur människors liv, hälsa, sociala och ekonomiska levnadsvillkor formats och förändrats i samspel med samhällets utveckling under mer än 200 år. Genom länkning till andra forskningsregister möjliggörs även studier av flera aktuella medicinska forskningsfrågor. Våren 2021 lanserades ett egenutvecklat uttagsverktyg som gör det möjligt för forskare att kostnadsfritt ladda ned data kostnadsfritt från web-portalen swedpop.se. Data som innehåller personuppgifter hanteras inom ramen för gällande lagstiftning och lämnas ut efter föreskriven etik- och sekretessprövning.

Kärnan i SwedPop utgörs av fem stora forskningsdatabaser: Databaserna POPUM och POPLINK vid Demografiska databasen (DDB), Umeå universitet, Skånes Ekonomisk-Demografiska Databas (SEDD) vid Centrum för ekonomisk demografi i Lund, Rotemansdatabasen vid Stockholms stadsarkiv, SweCens-databasen vid Riksarkivet och Gothenburg Population Panel (GOPP) vid Institutionen för ekonomi och samhälle, Göteborgs universitet. De storskaliga databaserna DDB och SEDD innehåller länkade individdata baserade på den svenska kyrkobokföringen i utvalda regioner, som gör det möjligt att följa individer, familjer och hushåll i sin sociala och ekonomiska kontext från sent 1600-tal till 1960-talet. Rotemansdatabasen innehåller motsvarande data från Stockholm 1877-1926. SweCens-databasen baseras på de svenska historiska folkräkningarna med information om hela den svenska befolkningen och GOPP är en panelstudie baserad på folkbokförings- och mantalsskrivningsdata från Göteborg under 1900-talet. Tillsammans erbjuder de fem databaserna sammanhängande information om den svenska befolkningen under nära tre sekler, med regionala nedslag i skilda delar av Sverige. Data omfattar såväl stad som landsbygd samt olika socioekonomiska miljöer. Det ökar möjligheterna att göra jämförande studier, till exempel av samhälls- och befolkningsutveckling i olika svenska regioner, i städer och på landsbygden. Även andra databaser kan knytas till SwedPop för att öka infrastrukturens vetenskapliga värde och täckningsgrad. En sådan databas är Sveriges Dödbok med information om avlidna svenskar, insamlad av Sveriges Släktforskarförbund. Sveriges Dödbok inkluderas i SwedPop mot slutet av bidragsperioden 2018-2022. Databasen förvaltas av Riksarkivet, genom avtal med Sveriges Släktforskarförbund. Folkräkningen 1950 har även förvärvats för projektets räkning och förvaltas av CEDAR/DDB

Övergripande målsättningar för projektet

Några av databaserna i SwedPop är redan väletablerade i forskningsvärlden, medan andra inte har använts lika mycket för forskning. För att öka forskningspotentialen, för samtliga databaser och för infrastrukturen som helhet, kommer utvecklingen av SwedPop under bidragsperioden 2018-2022 att fokusera på följande övergripande områden:

- Samordning av av kodning av yrken, dödsorsaker och geografiska enheter i de ingående databaserna.
- Strukturering av data enligt en gemensam databasstruktur med enhetlig nomenklatur och kodningssystem.
- Tillgängliggörande av data för forskare från en gemensam portal, swedpop.se
- Länkning av individer mellan de olika databaserna.
- Insamling av kompletterande data för att fylla luckor i tid och rum samt för att öka jämförbarheten mellan databaserna.
- Etablering av SwedPop som nationellt centrum för äldre befolkningsdata (före SCB-registren) på mikronivå.

Särskilda villkor för bidraget (2017-00666) samt budget och övergripande arbetsplan fastställdes av Vetenskapsrådet 2017-12-21.

I september 2021 beviljades SwedPop fortsatt stöd från Vetenskapsrådet för perioden 2023-2026. Särskilda villkor för bidraget (2021-00183) kommer att fastställas under våren 2022.

Utgångspunkter och struktur

Verksamhetsberättelsen för SwedPop följer samma struktur som projektets verksamhetsplan och tar sin utgångspunkt i de moduler som definieras i av VR fastställd budget och övergripande arbetsplan för projektet under bidragsperioden. För arbetet inom varje modul finns uppsatta delmål, vilka sträcker sig över kortare eller längre tidsperioder. Arbetet inom respektive modul leder också fram till ett antal mätbara resultat, s.k. delmål, vilka utgörs av färdigställda verktyg, datainsamlingar, tillgängliggörande av data osv. Dessa delmål definieras och beskrivs i den strategiska planen för SwedPop. I verksamhetsberättelsen kommer dessa delmål och övriga resultat att diskuteras under respektive modul.

2. Verksamheten under 2021

Organisation

Infrastrukturens styrgrupp har under året haft följande sammansättning:

Professor Dieter Müller	Umeå universitet	Ordförande
Professor Susanna Fellman	Göteborgs universitet	
Docent Therese Nilsson	Lunds universitet	
Avdelningschef Pauline Tjergefors	Riksarkivet	
Utvecklingschef Gabriel Marawgeh	Stockholms stadsarkiv	

Infrastrukturens ledningsgrupp har under året haft följande sammansättning:

Elisabeth Engberg	Umeå universitet	Föreståndare och sammankallande
Christer Lundh	Göteborgs universitet	
Martin Dribe	Lunds universitet	
Anders Nordström	Riksarkivet	
Mats Berggren	Riksarkivet	
Johan Gidlöf	Stockholms stadsarkiv	
Maria Larsson	Umeå universitet	

Styrgruppen har under 2021 sammanträtt vid två tillfällen och ledningsgruppen har haft sex protokollförda sammanträden, varav fem digitala.

Revision av verksamheten har genomförts, utan anmärkningar.

För att leda det operativa arbetet i de gemensamma modulerna M2-M4 har arbetsgrupper tillsatts med representation av samtliga parter i konsortiet.

Verksamhet allmänt

Verksamheten under året har bedrivits i enlighet med upprättad verksamhetsplan, strategisk plan och av VR fastställd budget för de olika modulerna. För att skapa förutsättningar för en god ekonomistyrning av projektet har en särskild ekonomibilaga utarbetats där parternas bidrag och medfinansiering tydliggörs för varje modul. Projektets organisation och samarbetet mellan konsortiets parter regleras i ett särskilt samarbetsavtal (dnr: CEDAR-5.B1-26-18)

Verksamhetsberättelsen fastställdes 2022-04-25 av projektets styrgrupp.

Arbetet under 2021 har i huvudsak följt upprättad verksamhetsplan och uppsatta budgetramar. Dock har effekterna av den pågående covid-19-pandemin fortsatt påverkat och försenat arbetet i flera moduler.

- Rekryteringen av en referensgrupp med sakkunskap om dödsorsaker (M2) har skjutits på framtiden.
- I M5 har registrering av uppgifter från digitalt tillgängliga källor fått prioriteras då tillgängligheten till fysiskt källmaterial har reducerats under pandemin. Omprioriteringen gör att delmål 2021:a inte har kunnat uppnås under året och kvarstår under 2022.
- Personalbrist och nödvändiga omprioriteringar vid Stockholms stadsarkiv under pandemin har orsakat kvarstående förseningar i M8. Förseningarna i M6-7 har arbetats in under 2021. Intentionen är att återstående moment kan slutföras under 2022.
- I M11 och M12 har arbetet försenats till följd av omprioriteringar när stora delar av personalen har arbetat hemifrån under pandemin. Det kvarstående arbetet i M11 kunde slutföras under hösten 2021.

Även om arbetet har påverkats av den pågående pandemin, räknar vi ändå med att kunna slutföra våra åtaganden inom projektets löptid, om än med vissa förseningar. Arbetet inom

M6, 7, 11 och 14 är nu avslutat. Vi kan också konstatera att arbetet med harmonisering av variabler och datastruktur (M2-3) nu är genomfört. Viss kodning återstår.

Avseende den ekonomiska redovisningen kan följande kommenteras:

- I M1 har medelsförbrukningen under 2021 varit avsevärt lägre än budgeterat. Detta beror på minskade kostnader för resor och logi under covid-19-pandemin, när fysiska möten inom projektet har ersatts av digitala möten.
- I M2 har Riksarkivets kostnader fortsatt varit högre än budgeterat, beroende på att arbetet med harmonisering har tagit längre tid och varit mer omfattande än planerat.
- I M3 har delar av arbetet som budgeterats för 2021 (Riksarkivet) skjutits fram till 2022.
- I M4 finns kostnadseftersläpningar beroende på omprioriteringar av arbetet inom modulen.
- Medelsförbrukningen i M5 har varit något lägre än förväntat, till följd av covid-19-pandemin och ej förutsedd sjukfrånvaro.
- Kostnadseftersläpningarna i M8 kvarstår. Personalbrist under den pågående pandemin har gjort att de inte har kunnat regleras som planerat under 2021.
- I M9 och M10 har kostnader tidigare lagts, till följd av att personalförändringar kommer att ske under 2022.
- I M11 och M12 har medelsförbrukningen varit lägre än budgeterat, då delar av arbetet inom dessa moduler har fått sättas på paus under pandemin.
- I M13 har varit medelsförbrukningen avsiktligt varit högre än budgeterat. Takten för inregistrering har skalats upp med syfte att reglera kostnadseftersläpningar från tidigare år.
- Medelsförbrukningen i M14 har under året varit högre än budgeterat, vilket beror på att arbetet inte kunde starta förrän 2020, till följd av yttre omständigheter utanför projektets kontroll.
- Arbetet i M15 startade ett halvår senare än beräknat vilket har gjort medelsförbrukningen lägre än budgeterat.

Kommunikation och information

Under 2021 har infrastrukturens webbportal swedpop.se omarbetats och uppdaterats. Uttagsverktyg och data har publicerats och metadata har tillgängliggjorts tillsammans med dokumentation och information om databaser och data. En nyhetssida har lagts till där information om kommande releaser har publicerats.

Även under 2021 har medverkan vid konferenser och nätverksmöten befunnit sig på en låg nivå till följd av pandemin.

3. Moduler - måluppfyllelse

Modul 1

Organisation och ledning

Modul 1 har en stödjande roll i projektet. Här finns resurser för projektledning, administration, resor och projektmöten. Inga specifika verksamhetsmål har satts upp för denna modul. Under 2021 har projektets kostnader för resor och logi varit lägre än budgeterat, till följd av den pågående pandemin. I övrigt följer verksamheten upprättad budget.

Modul 2

Harmonisering av kodning:

1. Harmonisering av variabler

Under 2021 har arbetet inom modul 2 fokuserat på standardisering och implementering av fastställda kodscheman, samt kodning av tillkommande data. När nya data läggs till i databaserna innebär det i de allra flesta fall också att fastställda kodscheman behöver kompletteras med ny information. Till exempel har standardiseringen och kodningen av 1950 års folkräkning inneburit att en rad nya yrke har tillkommit i kodschemat för yrkeskodning. Rutiner för det löpande arbetet med implementering och harmonisering av nya koder i SwedPop finns fastställda.

1a: Dödsorsaker

Det grundläggande arbetet med harmonisering av dödsorsakskodning i de ingående databaserna till SwedPop-standard är avslutat och kodschema samt rutiner för löpande kodning av nya dödsorsaker till har tagits fram. Kodschema och dokumentation av fastställda principer för kodning av dödsorsaker finns sedan 2021 öppet tillgängliga på [swedpop.se https://swedpop.se/data-description/](https://swedpop.se/data-description/). Till följd av covid-19-pandemin har ambitionen att utse en referensgrupp för kodning av nya dödsorsaker tillfälligt lagts på is.

Det faktiska arbetet med omkodning av dödsorsaker till SwedPop-standard i Umeå-, Lund- och Rotemansdatabaserna har fortsatt under 2021. SweCens- och GOPP-databaserna innehåller inga dödsorsaker.

1b. Yrken

Det grundläggande arbetet med harmonisering av yrkeskodning till SwedPop-standard har slutförts. Ett kodschema för yrkeskodning, *Yrkeskodning i SwedPop* har tagits fram och rutiner för fastställande av nya yrkeskoder, när behov av sådana uppstår under det löpande arbetet, har tagits fram. Under 2021 har även standardiseringen av samtliga yrkestitlar avslutats och i april fastställde ledningsgruppen en uppdaterad version av kodschemat, v.3, där samtliga yrken är både kodade och standardiserade. En ny version av kodschemat som också inkluderar kodade och standardiserade yrken i 1940 och 1950 års folkräkningar samt i Sveriges dödbok har också förberetts under året. Denna kommer att fastställas i februari 2022. Implementeringen av den nya yrkeskodningen (genomförs av varje databas för sig), där ca 1/3 av strängarna har fått nya koder har lett till att kodningen blivit mera precis.

Kodschema och dokumentation av de principer som används för yrkeskodning till SwedPop-standard finns sedan våren 2021 öppet tillgängliga på swedpop.se
<https://swedpop.se/data-description/>.

Avsikten är att all löpande yrkeskodning inom SwedPop i framtiden i hög grad ska kunna utföras maskinellt.

1c. Församlingskodning

Under 2021 har det grundläggande arbetet med harmonisering av församlingskodning avslutats. En gemensamt kodschema för kodning av församlingar till SwedPop-standard har fastställts, baserat på Riksarkivets register TORA *Topographic Register at the Swedish National Archives*. Då församlingsgränser inte sällan förändras över tid kan nya versioner av kodschemat behöva tas fram när nya data läggs till i databaserna. Den första uppdateringen kommer att behöva göras redan 2022 föranledd av att data från 1950-års folkräkning inkluderas i SwedPop. Dokumentation och kodschema finns sedan våren 2021 öppet tillgängliga på swedpop.se <https://swedpop.se/data-description/>.

Implementering av den geografiska kodningen utförs av varje databas för sig. Harmoniseringen av församlingskodning är därmed slutförd.

B. IDS-konvertering av databaserna:

Arbetet med harmoniseringen av databaserna i SwedPop till en gemensam databasstruktur, det internationella formatet IDS - *Intermediate Data Structure*, startade redan under samordningsprojektet. Efter en omfattande probleminventering och analys kunde gemensamma definitioner och principer för konverteringen till IDS 4.0 fastställas under 2020. Under våren 2021 gjordes en detaljerad dokumentation inklusive metadata avseende konvertering av data till SwedPop:s gemensamma databas format öppet tillgänglig på swedpop.se, <https://swedpop.se/ids-structure/>. Därmed har det grundläggande arbetet med harmonisering till IDS-format slutförts.

Varje part ansvarar för att konvertera sina data till fastställt format, samt för att ta fram erforderlig systemdokumentation för konverteringen till IDS 4.0. Inför den första datareleasen på swedpop.se i juni 2021 genomfördes noggranna tester för att säkerställa fullständig harmonisering mellan de ingående databaserna.

Modul 3

A. Utveckling av uttagsverktyg på webportalen

Arbetet startade 2018 och fortsatte under 2019 med framtagande av en detaljerad kravspecifikation för uttagsverktyget. Systemutvecklingen har ägt rum vid DDB och SEDD och Riksarkivet har deltagit i planeringen av det pågående arbetet. Utvecklingen av uttagsverktyget visade sig ta mer tid i anspråk än beräknat, då det tidigt stod klart att den underliggande databasstrukturen behövde kompletteras för att nå önskvärd funktionalitet. En reviderad projektplan fastställdes av ledningsgruppen 2019-10-29 och arbetet med utveckling av verktyget fortsatte under 2020-2021. En testversion togs fram 2020, som vidareutvecklades efter användartester. I juni 2021 kunde uttagsverktyget lanseras på swedpop.se där forskare nu kan ladda ned harmoniserade data från DDB, SEDD och SweCens (folkräkningen 1880), som inte utgör personuppgifter. Därmed har delmål 2020:a uppfyllts. Data från övriga databaser kommer att tillgängliggöras så snart kompletteringar, kodning och harmonisering är klar.

B. Utveckling av webbportal

Under 2021 har swedpop.se omarbetats och uppdaterats. Utagsverktyg och data har publicerats och metadata har tillgängliggjorts tillsammans med dokumentation och information om databaser och data. Publicerade data har försetts med versionshantering och PID kopplat till varje version.

Modul 4

Utveckling av verktyg, för att länka samman databaserna i SwedPop

Länkingsarbetet kan delas in i tre steg:

1. Länkning av samtliga SwedPop-databaser till Sveriges Dödbok.
2. Länkning av SEDD, DDB, Rotemannen och GOPP till folkräkningarna i SweCens.
3. Länkning av folkräkningarna till varandra.

Under första delen av bidragsperioden fokuserade arbetet på att testa och analysera metoder för att länka folkräkningar och kyrkboksdata (SEDD, DDB) till Sveriges dödbok. Testlänkningar har utförts och analyserats och förberedande arbete pågår inför utveckling av ett gemensamt verktyg för länkning av databaserna till den nya uppdaterade versionen av

Sveriges dödbok (se modul 15) med harmoniserad församlingskodning enligt SwedPop-standard. Tester har också visat på behovet av standardisering av förnamn och efternamn inom samtliga SwedPop databaser.

Under hösten 2021 genomfördes en workshop med fokus på principbeslut inför de tre länkingsstegen samt diskussioner kring testning steg 2: länkningen av samtliga databaser till folkräkningarna i SweCens samt steg 3: sammanlänkning av folkräkningarna till varandra. Där diskuterades även behovet av att utvärdera en ny länkingsmetod som utvecklats vid MPC i Minneapolis, baserad på machine learning. Metodens lämplighet för länkning av folkräkningar och för länkning mellan folkräkningar och kyrkboksdata i en svensk kontext kommer att testas och utvärderas.

Gemensamma principer och metoder/programvara kommer att tas fram för de ovan nämnda tre länkingsstegen. Därmed har delmål 2019:c behövt omformuleras. Det har också behövt senareläggas till 2022, då länkningen kan genomföras först när flertalet av de kompletteringar som pågår inom projektets moduler är avslutade (modul 6-8, 15).

Modul 5

Utvidga GOPP-databasen med nya data

Under 2021 har arbetet i modul 5 fortskridit enligt fastställd tidsplan, med den modifieringen att registrering av uppgifter från digitalt tillgängliga källor (församlingsböcker och personakter) har prioriterats medan registrering från papperskällor skjutits upp på grund av pandemin. Detta gör att delmål 2021:a inte har kunnat nås under året och kommer att färdigställas först under 2022. Alla urvalspersoner har länkats till databaserna Sveriges dödbok och Sveriges befolkning 1990, och för cirka 84% har uppgifter registrerats från personakterna. Excerpering ur församlingsböckerna har gjorts för 40% av urvalet och pågår. Excerpering av yrke och inkomst från inkomstlängderna har inletts.

Modul 6

Komplettering och länkning av data i Rotemansdatabasen

När databasen Rotemannen byggdes upp registrerades ca 40% av alla poster med en förenklad metod som innebar att viss information utelämnades. För att öka databasens forskningspotential har denna information behövt kompletteras och länkas. Förstudie, validering och dokumentation utfördes under 2018 och kompletteringsarbetet startade 2019. Det fortsatte under 2020-2021 och kommer att vara avslutat under 2022, försenat i förhållande till tidigare tidsplan. Förseningen är föranledd av Stadsarkivets flytt till nya lokaler 2019, vilket orsakade tidsmässiga omprioriteringar mellan modul 6-8. Under 2020 har arbetet försenats ytterligare till följd av personalbrist och förändrade prioriteringar under

den pågående covid-19-pandemin. Under 2021 har delar av standardiseringsarbetet samordnats med modul 15, så förseningar avseende kodning av yrke och geografi ([delmål 2021:b](#)) har kunnat arbetas in.

Modul 7

Komplettering av dödsorsaker i Rotemansdatabasen

Under 2021 har den kompletterande inregistreringen och länkningen av dödsorsaker från dödsorsaksintyg genomförts ([delmål 2020:e](#)). Arbetet skulle ha varit avslutat 2020, men har blivit försenat till följd av personalbrist och förändrade prioriteringar under den pågående covid-19-pandemin.

Modul 8

Komplettera Rotemansdatabasen med ekonomiska data

För att öka den komparativa forskningspotentialen i SwedPop ska Rotemansdatabasen kompletteras med ekonomiska data. Under 2019 och 2020 har ett förberedande arbete utförts med 1901 års huvudförteckning (den del av mantalsuppgifterna som listar samtliga lägenheter med information om innehavare, rumsantal, hyra, mm), som finns tillgänglig men i råformat och endast partiellt länkad. Avsikten var att därtill registrera in ekonomiska data från taxeringslängder för åren 1901 och 1910, men dessa planer fick revideras när projektiden reducerade från 8 till 5 år. Under 2021 har arbetet inletts med förberedande aktiviteter, skanning, systemutveckling och förberedelser för länkning av ekonomiska data. En fortsatt inregistrering av mer omfattande uppgifter om inkomst, skatt, hyra och fastighetsinnehav från mantalslängder under en längre tidsperiod 1880-1910, samt länkning kommer att ske under projektperioden 2023-2026.

Modul 9

Komplettering av SEDD med ekonomiska data för Landskrona 1905-1947

Arbetet inom denna modul avser en komplettering av ekonomiska data i form av inkomst- och taxeringslängder för tidsperioden 1905-1947, för att öka databasens forskningsvärde. Motsvarande data för tidsperioden 1948-1968 finns sedan tidigare inkluderade i databasen och används för forskning. Till och med 2021 har inkomstlängderna för åren 1906, 1910, 1912, 1915, 1920, 1923, 1925, 1928, 1930, 1933, 1935, 1938, 1940, 1942 och 1944-47 registrerats och länkats till databasen. Under 2019 justerades arbetsplanen för modulen av arbetsekonomiska skäl något i förhållande till projektets långsiktiga strategiska plan: Istället

för att registrera alla år i ordning, så registreras längderna först med tio års mellanrum, därefter inkluderas data för mellanliggande år ([delmål 2021:c](#)).

Modul 10

Komplettering av SEDD med befolkningsdata för Landskrona 1895-1910

Befolkningsdata för Landskrona 1910-1968, och från omkringliggande församlingar 1813-1968 har inkluderats i SEDD inom ramen för tidigare projekt. Under perioden 2018-2021 har information från församlingsböcker samt register över födda, döda, in- och utflyttade har registrerats för perioden 1905-1909 och länkats till övriga data i databasen. Databasen har även kompletterats med information om faders yrke för individer i Landskrona födda 1905-1947 ([delmål 2021:e](#)).

Modul 11

Komplettering av DDB med befolkningsdata för Nysätra församling

Arbetet med att inkludera befolkningsdata från Nysätra i DDB-databaserna startade 2019 och avslutades 2021. Därmed är [delmål 2021:d](#) uppfyllt. Ända fram till våren 2020 följde arbetet upprättad tidplan, men försenades till följd av den pågående covid-19-pandemin och övergången till hemarbete. De avslutande länkningsmomenten som kräver tillgång till den skyddade datormiljön på arbetsplatsen kunde slutföras först under hösten 2021. Data tillgängliggjordes i databasen POPLINK i december 2021 och kommer att tillgängliggöras i SwedPop:s uttagsverktyg på swedpop.se under våren 2022.

Modul 12

Komplettering av DDB med ekonomiska data för Lövångers församling

För att öka den komparativa forskningspotentialen i SwedPop kompletteras DDB:s befolkningsdatabaser med ekonomiska data. Arbetet startade under 2021, försenat till följd av den pågående covid-19-pandemin.

Modul 13

Inkludering av 1940 års folkräkning i SwedPop

Arbetet startade under 2018. Vissa arbetsmoment tidigarelades, vilket gjorde att arbetet vid 2019 års ingång låg något före fastställd projektplan. I samarbete med Sveriges Släktforskarförbund startade under hösten 2019 ett storskaligt citizen- science-projekt där frivilliga släktforskare deltar i inregistreringsarbetet. Under 2020-2021 har registreringsarbetet skalats upp i takt med att allt flera frivilliga släktforskare anslutit sig till projektet. Funktioner och formuläret i inregistreringsprogrammet och tillhörande manualer har också förbättrats under året, för att effektivisera inregistreringsarbetet. Datainsamlingen följer hittills upprättad plan, men förhoppningen är att samarbetet med Sveriges Släktforskarförbund ska bidra till att fler församlingar än planerat kan inkluderas under projekttiden. Arbetet med volontärer har gjort att den ursprungliga tanken med att färdigställa folkräkningen län för län har behövt omprövas. En del i citizen-science-strategin är att volontärerna själva väljer vilka län de vill arbeta med, vilket gör att arbetet blir mer splittrat. I nuläget är prognosen att Jämtland och Värmland ligger närmast till att bli klara under 2022. Detta gör att delmål 2022:h delvis har behövt omformuleras.

Modul 14

Inkludering av 1950 års folkräkning i SweCens-databasen

Under 2021 avslutades arbetet med kodning och upparbetning av 1950 års folkräkning, vilket innebär att variabler har harmoniserats och kodats till SwedPop-standard (delmål 2020:g).

Modul 15

Sveriges dödbok

Sveriges dödbok 1860-2017 är en databas som innehåller information om alla individer som har avlidit i Sverige från 1860 och fram till idag. Genom ett avtal med Sveriges Släktforskarförbund har SwedPop fått tillgång till dessa data för att inkludera dem som en del i SwedPop. Databasen är en väsentlig resurs för länken mellan olika databaser i SwedPop och, då poster för avlidna efter 1947 innehåller personnummer som gör det möjligt att koppla SwedPop-material, till andra externa databaser, exempelvis från SCB eller Skatteverket. Huvuddelen av databasen har skapats av frivilliga släktforskare, med målet att sökbarheten för genealogisk och kvalitativ forskning ska vara så bra som möjligt. Detta innebär att råmaterialet har vissa brister i standardisering och kodstruktur vid

användning för kvantitativ forskning och för länkning till forskningsdatabaser. Under 2020 fick Sveriges Släktforskarförbund i uppdrag av SwedPop-konsortiet att identifiera och minimera dessa kvalitetsbrister. Arbetet avslutades under 2021 och har omfattat nyregistrering av födelseuppgifter, validering och korrigering av namnuppgifter, kön och födelseort, kontroll av personnummer och kontroll av att myndighetsuppgifter om dödsfall stämmer, att personerna verkligen är döda. Det senare har betydelse för om uppgiften kan klassificeras som personuppgift eller inte enligt GDPR. Under 2021 har det också genomförts en komplettering med döda 2018-2020 och namn, orter och yrken har harmoniserats och kodats till SwedPop-standard.

4. Tillgängliggörande av data

SwedPop är en forskningsinfrastruktur under uppbyggnad vilket givetvis påverkar förutsättningarna för tillgängliggörandet av data. Den första releasen av harmoniserade forskningsdata gjordes den 29 juni 2021 och omfattade en begränsad mängd data från DDB, SEDD och SweCens (1880 års folkräkning). En ny release planeras under våren 2022. En betydligt större mängd data från de databaser som ingår i SwedPop har tillgängliggjorts för forskningsändamål under året. Data från SweCens-databasen ingår också i det internationella IPUMS-International, där forskare fritt kan ladda ned dataset för forskning (<https://international.ipums.org/international/>). En översikt över antalet levererade dataset, antalet projekt och användare finns i den särskilda redovisningen av nyckeltal.

Under 2021 har data från DDB, Riksarkivet och Stockholms stadsarkiv även tillgängliggjorts i sökbara databaser på webben och på andra datamedier. Rotemansdatabasen (<http://digitalastadsarkivet.stockholm.se/Rotemannen2012/Search.aspx>) och data från svenska folkräkningar (<https://sok.riksarkivet.se/folkrakningar>) finns fritt tillgängliga på webben och DDB-data fram till 1900 finns sökbara i applikationen Indiko-kyrkböcker på nätet (<https://www.umu.se/enheten-for-demografi-och-aldrandeforskning/sokverktyg/indiko/>). Ingen statistik finns på antalet användare, då databaserna är öppett tillgängliga utan inloggning. Dessa sökapplikationer är i första hand anpassade till en intresserad allmänhet, men används i vissa fall även av forskare trots att möjligheten finns att begära mer komplexa uttag för forskningsprojekt. Detsamma gäller databasen Sveriges Dödbok, som inte har funnits på webben utan enbart som fristående sökdatas, men där Sveriges Släktforskarförbund har godkänt och förmedlat mer avancerade forskningsuttag.