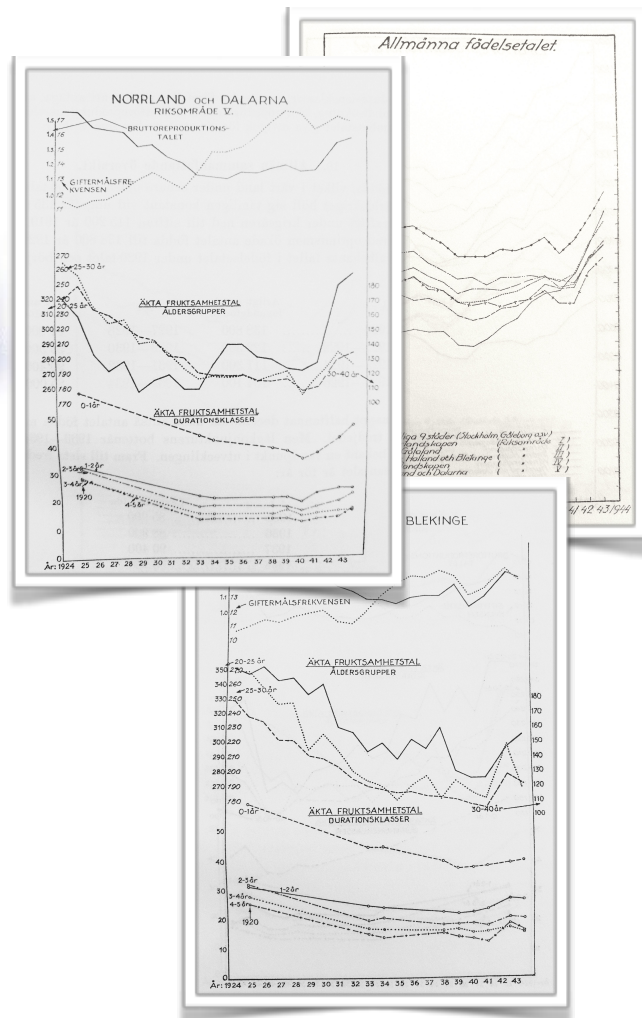




SwedPop finansieras av Vetenskapsrådet

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.
- **Stockholms stadsarkiv: Rotemansarkivet** Livsförloppsdata för Stockholms befolkning 1877-1926.

SwedPop: Svenska befolkningsdatabaser för forskning 2018-2022

1. Inledning

Bakgrund

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. För att öka tillgängligheten till data kommer forskare att kunna 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 och olika socioekonomiska miljöer, vilket 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 i förlängningen komma att knytas till SwedPop för att öka infrastrukturens vetenskapliga värde och täckningsgrad. En av dessa är Sveriges Dödbok med information om avlidna svenskar, insamlad av Sveriges Släktforskarförbund, som inkluderas i SwedPop mot slutet av bidragsperioden 2018-2022. Databasen kommer att 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 att fokusera på följande övergripande områden:

- Samordning av av kodning av yrken, dödsorsaker och geografiska enheter i den ingående databaserna
- Strukturering av data enligt en gemensam databasstruktur (IDS) med enhetlig nomenklatur och kodningssystem
- Tillgängliggörande av data för forskare från 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.

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, som utgörs av färdigställda verktyg, datainsamlingar, tillgängliggörande av data osv. (beskrivs vidare i den strategiska planen för SwedPop). Delmål och resultat uppnådda under 2020 kommer att diskuteras under respektive modul.

2. Verksamheten under 2020

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	
Forskningssekreterare Cecilia Notini	Riksarkivet t.o.m. 2020-07-15	
Avdelningschef Pauline Tjergefors	Riksarkivet fr.o.m 2020-07-16	
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 2020 sammanträtt vid två tillfällen och ledningsgruppen har haft sju protokollförda sammanträden.

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 även detta år 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)

Arbetet under 2020 har med följande undantag följt upprättad verksamhetsplan och uppsatta budgetramar:

- Effekter av den pågående Corona-pandemin har påverkat och försenat arbetet i flera moduler.
 - Rekryteringen av en referensgrupp med sakkunskap om dödsorsaker (M2) har skjutits på framtiden.
 - Det pågående arbetet i M6-M7 har försenats ytterligare något till följd av personalbrist och behov av omprioriteringar vid Stockholms stadsarkiv under den pågående pandemin. Avsatta medel flyttas till 2021 och bedömningen är att arbetet kommer att kunna slutföras inom projekttiden, om än något försenat.
 - Länkningen av insamlade data från Nysätra församling (M11) pausades under våren 2020 i samband med övergång till hemarbete för medarbetare vid Umeå universitet. Detta gör att M11 kommer att färdigställas först 2021.
- Arbetet i M2 med IDS-konvertering av samtliga databaser har blivit något försenat. Detta beror på dels på att det visade sig vara mer komplicerat än tidigare förutsett, dels på att det har behövt samordnas med det pågående arbetet med uttagsverktyg i M3.
- Arbetet i M3 följer den reviderade projektplan som antogs av ledningsgruppen 2019-10-29 och vid 2020 års utgång hade en betaversion av uttagsverktyget utvecklats, enligt upprättad plan. Däremot har publiceringen av uttagsverktyget på swedpop.se fördröjts något i väntan på testning och på leverans av IDS-konverterade och harmoniserade data. Utagsverktyget kommer att publiceras under våren 2021.
- I M4 har, som tidigare diskuterats, metodologiskt motiverade omprioriteringar gjorts under arbetets gång, vilket gör att den ursprungliga tidsplanen har behövt revideras.

Sammantaget utgör dessa avsteg från verksamhetsplanen smärre kursändringar och tidsmässiga förskjutningar som inte kommer att få några omfattande konsekvenser för projektet i ett längre perspektiv. Även om arbetet under 2020 har påverkats av den pågående Corona-pandemin, räknar vi ändå med att kunna slutföra våra åtaganden inom projektets löptid. Kostnadseftersläpningar kommer att uppstå, men sett i ett femårsperspektiv kommer förseningarna inte att ha någon budgetpåverkan.

En viktig milstolpe under 2020 har varit att arbetet i M14 äntligen har kunnat starta, två år försenat i förhållande till tidigare upprättad projektplan. Det kommer att slutföras under våren 2021. För att klara inköpet av data från Arkiv Digital inom befintliga budgetramar har tidigare avsatta medel i M14 avsedda för kodning och upparbetning av data behövt tas i anspråk. Detta har kompensats av en ökad medfinansiering från UmU och Riksarkivet.

Avseende den ekonomiska redovisningen kan följande kommenteras:

- I M1 har medelsförbrukningen varit lägre än budgeterat beroende på radikalt minskade kostnader för resor och logi under Corona-pandemin.
- I M2 har Riksarkivets kostnader varit högre än budgeterat, beroende på att arbete som budgeterats 2018-2019 har utförts först under 2020.
- Kostnadseftersläpningarna i M6-8 till följd av Stockholms stadsarkivs flytt till nya lokaler och tekniska omprioriteringar mellan modulerna kvarstår. Avsikten var att dessa skulle regleras under 2020, men detta har inte varit möjligt till följd av den pågående pandemin. Det innebär att avsatta medel kommer att förbrukas först under 2021 och 2022.
- En kostnadseftersläpning finns i M13. Medlen kommer att förbrukas under 2021, när inregistreringen skalas upp ytterligare.
- I M14 har medel förts över från RA till medelsförvaltaren UmU, som har genomfört upphandlingen av data.
- Kostnadseftersläpningar finns även i M15, där arbetet startade ett halvår senare än beräknat.

Kommunikation och information

Det fortsatta arbetet med webbportalen **swedpop.se** har legat på is under 2020, delvis till följd av brist på informatörsresurser. I samband med lanseringen av uttagsverktyget tidigt under 2021 kommer en större omarbetning att göras, i syfte att utveckla portalen till att bli mer användarvänlig. Flera planerade informationstillfällen har fått ställas in, då i stort sett all konferensverksamhet har legat nere under 2020, till följd av pandemin. Avsikten var bland annat att SwedPop och forskning på SwedPop-data skulle presenteras vid en session på European Social Science History Conference i Leiden i mars och vid en workshop med det europeiska SHIP-projektet i Edinburgh i maj. Däremot har de nordiska kontakterna utvecklats under året. I februari presenterades SwedPop vid det danska systerprojektet LinkLives årsseminarium i Köpenhamn. En digital presentation har också hållits med företrädare för det nationella projektet Norwegian Historical Population Register (HPR) i Oslo.

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 2020 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 2020 har arbetet inom modul 2 övergått i nästa fas. Den grundläggande harmoniseringen av variablerna dödsorsak, yrke och församling utfördes under 2018-2019 och kodscheman har tagits fram och fastställts. Från 2020 och framåt ligger fokus på standardisering och implementering samt kodning av tillkommande data. När nya data läggs till i databaserna innebär det i regel att fastställda kodscheman behöver uppdateras och 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 har tagits fram och fastställts.

1a: Dödsorsaker

Det grundläggande arbetet med harmonisering av dödsorsakskodning avslutades under 2019. Detta har under 2020 kompletterats med harmonisering av tillkommande dödsorsaker i Rotemannen, som av arbetsekonomiska skäl har synkroniserats med arbetet i M7. Fortsatt harmonisering av dödsorsaker sker löpande, när nya data läggs till. Rutiner för löpande kodning av nya dödsorsaker i de ingående databaserna togs fram under 2019. En dokumentation av processen har även utarbetats och kommer att tillgängliggöras på swedpop.se i samband av lanseringen av uttagsverktyget. Avsikten var att en referensgrupp med sakkunskap om historiska dödsorsaker skulle utses under 2020, men till följd av den pågående Corona-pandemin har detta skjutits på framtiden. Under 2020 har arbetet inom det internationella projektet SHIP, där SwedPop är representerat, resulterat i en första version av en ny internationell standard för dödsorsaker historiska data, ICD10H. Avsikten är att den i framtiden ska implementeras som en ny standard i SwedPop.

1b. Yrken

Under 2020 har arbetet fortsatt med standardisering och implementering av SwedPop - HISCO. En uppdaterad version av kodschemat, SwedPop-HISCO 2.0, fastställdes i juni 2020, med tillhörande dokumentation. Därmed har en svensk standard för HISCO etablerats, vilket inte bara kommer att gynna detta projekt. Arbetet har fortsatt med standardisering, kodning och implementering enligt fastställda rutiner.

1c. Församlingskodning

Arbetet med harmonisering av församlingskodning återupptogs under 2020, föranlett av att nya data tillförts SwedPop. Under 2020 fick projektet tillgång till folkräkningen 1950 (M14) med uppgifter om boendeförsamlingar, födelseförsamlingar och födelseländer för sju miljoner personer, vilket innebär ett stort tillskott av geografisk information att analysera. Detta gjordes under hösten 2020, vilket ledde till en ytterligare revidering av kodschemat för församlingar, samt en kodning av födelseländer baserat på kodschemat från IPUMS vid University of Minnesota. Arbetet med harmonisering av församlingskoder kommer att fortsätta under 2021. Efter analysen av folkräkningen 1950 vidtar analysen av den senaste versionen av Sveriges dödbok (modul 15). Arbetet med församlingskodningen kommer att behöva pågå till dess analysen av Sveriges Dödbok är genomförd, vilket beräknas ske under våren 2021.

B. IDS-konvertering av databaserna:

Det omfattande arbetet med att konvertera databaserna i SwedPop till ett gemensamt databasformat har fortsatt under 2020 under ledning av en särskild arbetsgrupp, utsedd av ledningsgruppen. Under 2020 har arbetsgruppen slutfört det tidvis komplicerade arbetet med att fastställa gemensamma definitioner och principer för konverteringsarbetet och en slutlig version av SwedPop-IDS har tagits fram tillsammans med metadata och systemdokumentation. Varje organisation i konsortiet arbetar vidare med att konvertera sina data till fastställt IDS-format. Arbetet är något försenat i förhållande till tidigare upprättad tidsplan då det delvis är avhängigt av och har behövt samordnas med det pågående arbetet i M3, se nedan.

Modul 3

A.Utveckling av uttagsverktyg på webportalen

Arbetet startade redan under 2018 med att ta fram riktlinjer för verktygets funktionalitet och fortsatte under 2019 med framtagande av en detaljerad kravspecifikation. Diskussionerna om kravspecifikation, arbetssätt, databasmodell och funktionalitet tog längre tid i anspråk än förutsett, vilket gjorde att de initiala planerna på en tidig lansering av ett enkelt uttagsverktyg fick överges. Fokus riktades istället mot att lansera ett fullt funktionellt uttagsverktyg under 2020 och en reviderad projektplan fastställdes av

ledningsgruppen 2019-10-29. Under 2020 har utvecklingsarbetet fortsatt i enlighet med upprättad plan och innan årsskiftet fanns en betaversion av uttagsverktyget tillgängligt för testning (**Delmål 2020:a**). Däremot har publiceringen av uttagsverktyget på swedpop.se fördröjts något i väntan på testning och på leverans av IDS-konverterade och harmoniserade data. Förseningen beror dels på arbetets komplexitet, dels på att det är avhängigt av och har behövt samordnas med IDS-konverteringen i M2, se ovan.

B. Utveckling av webbportal

Under 2020 har arbetet med att utveckla webbsidan inte prioriterats, i väntan på det uttagsverktyg som är under utveckling. Arbetet med att utveckla en nyhetssida har också skjutits fram till 2021 och den kommer att färdigställas lagom till lanseringen av uttagsverktyget.

Modul 4

Utveckling av verktyg, för att länka samman databaserna i SwedPop

Arbetet startade hösten 2018 och har fortsatt under 2019 och 2020. Målet är att i steg ett ta fram ett verktyg för att länka folkräkningarna till Sveriges Dödbok. I steg två utvecklas verktyg och metoder för att länka övriga databaser till Sveriges Dödbok och i ett tredje steg ska övriga databaser länkas mot folkräkningarna. Under 2020 fick projektet tillgång till folkräkningen 1950 vilket ledde till nya prioriteringar. En testlänkning av folkräkningen 1950 till Sveriges dödbok genomfördes och för att utforska möjliga länkningsprinciper testlänkades även SEDD till folkräkningen 1950. Slutgiltiga länkringar har inte kunnat genomföras ännu eftersom det dels krävs en genomförd harmonisering av födelseförsamlingar i samtliga SwedPop-databaser (modul 2) och dels en ny uppdaterad version av Sveriges Dödbok (modul 15). Arbetet med länkning har också visat på behovet av att även standardisera förnamn och efternamn inom samtliga SwedPop databaser. Det beräknas vara möjligt att genomföra kompletta skarpa länkringar av SwedPop-databaserna under hösten 2021. Uppfyllelsen av delmål 2020:b, utveckling av verktyg för att länka samman databaserna kommer att dröja till 2021, till följd av metodologiskt motiverade förändringar av projektplanen.

Modul 5

Utvidga GOPP-databasen med nya data

GOPP-databasen är en panelstudie som idag omfattar ett 1%-sample av befolkningen i Göteborg 1915-1943. I denna modul kommer databasen att utvidgas, genom att de personer som kan observeras 1943 följs fram till 1967 och att nya urvalspersoner födda

eller inflyttade efter 1943 tillförs och följs till 1967. Under 2020 har arbetet i modulen fortskridit enligt fastställd tidsplan. I nuläget har alla urvalspersoner länkats till databaserna Sveriges dödbok och Sveriges befolkning 1990, och för cirka hälften har uppgifter registrerats från personakterna. Excerpering ur församlingsböckerna har gjorts för 20% av urvalet och pågår.

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 behöver denna information kompletteras och länkas. Förstudie, validering och dokumentation utfördes under 2018 och kompletteringsarbetet startade 2019. Arbetet har fortsatt under 2020 och kommer att vara avslutat under 2022, försenat i förhållande till tidigare tidsplan. Förseningen är föranledd av Stadsarkivets flytt till nya lokaler 2019, vilken ledde till tidsmässiga omprioriteringar mellan modul 6-8. Under 2020 har arbetet försenats ytterligare till följd av personalbrist och förändrade prioriteringar under den pågående Corona-pandemin.

Modul 7

Komplettering av dödsorsaker i Rotemansdatabasen

För att Rotemansdatabasens information om dödsorsaker ska bli fullständig behöver en kompletterande inregistrering ske av uppgifter i Stadsläkarens arkiv. Under 2019 påbörjades inskanningen av samtliga dödsorsaksintyg och länkning av tidigare olänkade poster startade. En analys har också gjorts för att identifiera var kompletteringar behöver ske. Nyregistrering av dödsorsaker är nu genomförd men harmonisering, länkning och omkodning enligt SwedPop-standard återstår. Arbetet skulle ha varit avslutat 2020, men kommer att bli försenat. Förseningen är liksom i M6 föranledd av Stadsarkivets flytt till nya lokaler 2019, samt av personalbrist och förändrade prioriteringar under den pågående Corona-pandemin. Uppfyllelsen av delmål 2020:c kommer därmed att dröja till 2021.

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 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 2021-2022 kommer arbetet därför att fokuseras på förberedande aktiviteter, skanning, systemutveckling och förberedelser för länkning av ekonomiska data. Resurser för transkribering finns budgeterade för sista halvåret 2022. I ny ansökan till VR ansöker projektet om medel för nyregistrering av mer omfattande uppgifter om inkomst, skatt, hyra och fastighetsinnehav från mantalslängder under en längre tidsperiod 1880-1910, samt länkning.

Modul 9

Komplettering av SEDD med ekonomiska data för Landskrona 1905-1947

Arbetet inom modul 9 har löpt enligt upprättad tidplan. Det 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 2020 har inkomstlängderna för åren 1906, 1910, 1915, 1920, 1925, 1930, 1933, 1935, 1938, 1940, 1942 och 1944-47 registrerats och länkats till databasen (Delmål 2020:d). 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.

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. Arbetet inom modul 10 har löpt enligt tidplan. 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 2010:e).

Modul 11

Komplettering av DDB med befolkningsdata för Nysätra församling

Arbetet med att inkludera befolkningsdata från Nysätra som startade 2019 har fortsatt under 2020. Ända fram våren 2020 har arbetet följt upprättad tidplan, men har inte kunnat slutföras som planerat till följd av den pågående Corona-pandemin och övergången till hemarbete. Förseningen avser de avslutande länkningsmomenten som måste utföras i den skyddade datormiljön på arbetsplatsen. Arbetet är inte tidskritiskt i förhållande till övriga moduler och pausades därför under 2020. Det kommer att återupptas och avslutas under 2021.

Modul 12

Komplettering av DDB med ekonomiska data för Lövångers församling

Målet med M12 är att öka den komparativa forskningspotentialen i SwedPop genom att komplettera DDB:s befolkningsdatabaser med ekonomiska data, motsvarande de som finns tillgängliga i SEDD. Under 2020 har en förstudie av källmaterialet genomförts och ett inregistreringsformulär för datainsamling har utvecklats (Delmål 2020:f). Datainsamlingen kommer att starta redan 2021, före tidigare fastställd tidplan.

Modul 13

Inkludering av 1940 års folkräkning i SwedPop

Arbetet startade under 2018. Vissa arbetsmoment tidigare lades, vilket gjorde att arbetet vid 2019 års ingång låg något före fastställd projektplan. Ett nytt formulär i inregistreringssystemet PERSONA avsett för folkräkningar har tagits fram och driftsatts, inskanning av källmaterialet har genomförts och datainsamling har påbörjats. Vid 2020 års ingång hade de första församlingarna hunnit bli komplett inregistrerade. I samarbete med Sveriges Släktforskarförbund startade under hösten 2019 ett storskaligt citizen-science-projekt där frivilliga släktforskare deltar i inregistreringsarbetet. Under 2020 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.

Modul 14

Inkludering av 1950 års folkräkning i SweCens-databasen

Under 2020 kunde äntligen arbetet med att inkludera 1950 års folkräkning starta, kraftigt försenat till följd av omständigheter utanför projektets kontroll. Då det under 2019 stod klart att en ny överenskommelse mellan Riksarkivet och Arkiv Digital, som digitaliserat folkräkningen i genealogiskt syfte, inte skulle kunna nås förbereddes en upphandling av data för projektets räkning. Innan dess undersöktes även möjligheten att digitalisera data i egen regi, men källmaterialet bedömdes inte hålla en sådan kvalitet att OCR-läsning skulle vara möjlig. Tanken på en digitalisering med manuella metoder uteslöts helt, då en sådan varken skulle rymmas inom projektets tids- och budgetramar. Styrgruppen för SwedPop gav 2019-10-16 medelsförvaltaren mandat att gå vidare med en upphandling med målsättning att projektet skulle få tillgång till data innan 2019 års utgång. Upphandlingen genomfördes i enlighet med gällande lagstiftning, med Arkiv Digital som enda anbudsgivare och tilldelningsbeslut fattades i december 2019. Beslutet överklagades till Förvaltningsrätten av en annan aktör, vilket ytterligare fördröjde processen. Överklagan avslogs emellertid 2020-03-31 och 2020-04-15 kunde avtal slutligen träffas mellan medelsförvaltaren Umeå Universitet och Arkiv Digital om inköp av 1950 års folkräkning. Arbetet med kodning och upparbetning startade vid halvårsskiftet 2020 och kommer att vara avslutat under första halvåret 2021 (Delmål 2020:g). Inköpet kunde genomföras inom given budgetram, men krävde en viss omfördelning av resurser. Därtill har Umeå universitet och Riksarkivet gått in med ökad medfinansiering för att täcka kostnader för kodning och upparbetning.

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 för forskningsändamål fått tillgång till dessa data för att inkludera dem som en del i SwedPop. Databasen är en väsentlig resurs för länken mellan olika material 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. Dessutom saknas ibland fullständiga uppgifter om födelseförsamling, födelse- och dödsdatum. Sveriges Släktforskarförbund har under 2020, på uppdrag av SwedPop, arbetat med att identifiera och minimera dessa kvalitetsbrister. Arbetet har omfattat nyregistrering av födelseuppgifter, validering och korrigering av namnuppgifter,

genus 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 SwedPop-projektet har flera av de ingående databaserna provlänkats mot Sveriges dödbok, inom ramen för modul 4, och dessa erfarenheter är till stor nytta i det pågående förbättringsarbetet. Arbetet fortsätter och avslutas under 2021 med rättning och komplettering av uppgifter enligt ovan, harmonisering och kodning av namn, orter och yrken till SwedPop-standard samt komplettering med döda 2018-2020. Kodning till IDS-format utförs därefter av Riksarkivet.

4. Tillgängliggörande av data

SwedPop är en forskningsinfrastruktur under uppbyggnad vilket givetvis påverkar förutsättningarna för tillgängliggörandet av data. Inga dataset med harmoniserade data har ännu levererats från swedpop.se, då uttagsverktyget inte kommer att vara på plats förrän under första kvartalet 2021. Däremot har data från de databaser som ingår i SwedPop 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 2020 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 öppettillgängliga 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.