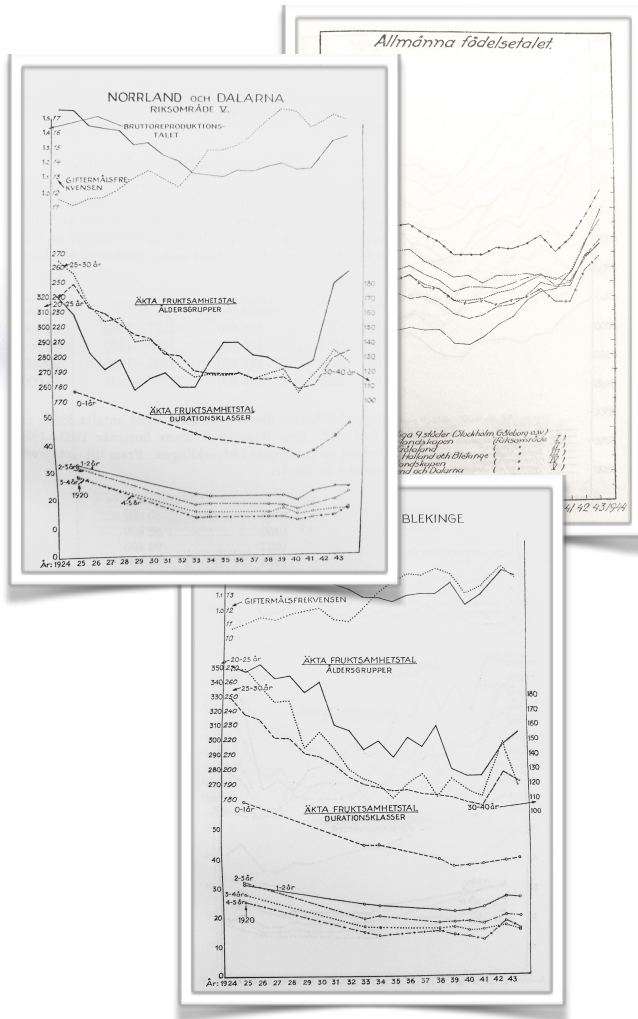


SwedPop

Svenska befolkningsdatabaser för forskning



UMEÅ
UNIVERSITET



GÖTEBORGS
UNIVERSITET



LUNDS
UNIVERSITET

RIKSARKIVET STOCKHOLMS STADSARKIV

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 personer från ca 100 församlingar i olika svenska regioner 1650-1960. Koordinator för projektet.
- **Göteborgs universitet: Göteborgs Befolkningspanel (GOPP)** Institutionen för ekonomi och samhälle. Fyraårspaneler för ett slumpvis urval bosatta i Göteborg under något av panelåren 1915-1943.
- **Lunds universitet: Skånes Ekonomisk-Demografiska Databas (SEDD)** Centrum för ekonomisk demografi. Livsförloppsdata för personer från ett antal skånska församlingar 1680-1968.
- **Riksarkivet: SweCens-databasen.** Folkräkningarna 1880-1950. Tvärsnitt vart tionde år för hela riket.
- **Stockholms stadsarkiv: Rotemansarkivet.** Livsförloppsdata för Stockholms befolkning 1877-1926.

1. Inledning

SwedPop: Svenska befolkningsdatabaser för forskning 2018-2022

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 kommer att samordnas och tillgängliggöras 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 över mer än 200 år. Genom länkning till andra forskningsregister möjliggörs därtill studier av flera aktuella medicinska forskningsfrågor.

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. DDB, SEDD och Rotemansdatabasen innehåller longitudinella individdata baserade på den svenska kyrkobokföringen i utvalda regioner. SweCens-databasen baseras på de svenska historiska folkräkningarna med information om hela den svenska befolkningen och GOPP är en panelstudie baserad på folkbokföringsdata från Göteborg under 1900-talet.

Tillsammans erbjuder de fem databaserna sammanhängande information om den svenska befolkningen från tidigt 1800-tal och fram till 1950, 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. I förlängningen kommer även andra databaser att kunna 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 kommer att inkluderas i SwedPop mot slutet av bidragsperioden 2018-2022.

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 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.

Utgångspunkter

Verksamhetsplanen för SwedPop tar sin utgångspunkt i de särskilda bidragsvillkor och den budget och arbetsplan som har fastställts av Vetenskapsrådet (2017-00666) för bidragsperioden 2018-2022.

Verksamhetsplan för varje år utarbetas av ledningsgruppen för SwedPop. Planen fastställs därefter av styrgruppen för projektet och ska under bidragsperioden vara Vetenskapsrådet tillhanda senast den 15 november.

Verksamhetsplanen är strukturerad utifrån de moduler som finns angivna i budget och arbetsplan för projektet. Planen fastställdes 2019-10-16 av styrgruppen för SwedPop.

Delresultat

Projektet är organiserat i moduler, där arbetet inom varje modul mynnar ut i ett antal mätbara delresultat (färdigställda verktyg, datainsamlingar, tillgängliggörande av data osv.). Planerade delresultat under bidragsperioden finns beskrivna i den strategiska planen för SwedPop fastställd av styrgruppen 2018-06-04.

Under 2020 är målsättningen att arbetet ska leda fram till följande delresultat:

A	Ett uttagsverktyg med full funktionalitet finns tillgängligt på SwedPop.se , där forskare kan hämta hem data från samtliga databaser.	3
B	Utveckling av verktyg för att länka samman databaserna i SwedPop är slutförd. *	4
C	Rotemansdatabasen har kompletterats med fullständig information om dödsorsak.	7
D	SEDD-databasen har kompletterats med inkomstdata för åren 1910, 1915 and 1920.**	9
E	SEDD-databasen har kompletterats med information om fars yrke för individer i Landskrona födda 1905-1947.	10
F	Inregistreringssystemet PERSONA har anpassats för digitalisering av inkomstuppgifter.	12
G	1950 års folkräkning har kodats och länkats till övriga folkräkningar, samt till övriga databaser i SwedPop.***	14

* Kvarstår från 2019. Se vidare text under modul 4.

**Av arbeteekonomiska skäl har ordningen för inregistreringen av längderna förändrats, vilket påverkar målformuleringen. Se vidare text under modul 9.

*** Kvarstår från 2019 till följd av av omständigheter utanför vår kontroll. Se vidare text under modul 14.

2. Moduler - aktiviteter och tidsplan

Modul 1

Organisation och ledning

Modul 1 har en stödjande roll i projektet. Här finns budget för projektledning, administration, resor och projektmöten. Inga specifika verksamhetsmål har utarbetats för 2020.

Modul 2

A. Harmonisering av kodning:

1. Harmonisering av variabler

- Dödsorsaker
- Yrke
- Församlingskodning

1a: Dödsorsaker

Nuläge

Kodningen av dödsorsaker påbörjades under samordningsprojektet 2016-2017 och har fortsatt under 2018 och 2019. Ingen särskild arbetsgrupp finns tillsatt för detta arbete. Dödsorsaker i Umeå och Lund är kodade enligt gemensamma principer och arbetet med harmonisering av dödsorsaker i Rotemansdatabasen kommer att vara genomförd under 2019. SweCens- och GOPP-databaserna innehåller inga dödsorsaker.

Under 2019 har rutiner för löpande kodning av nya dödsorsaker tagits fram. En löpande dialog förs med det internationella projektet SHIPS, där det för närvarande pågår ett arbete för att ta fram ett standardiserat kodningssystem för dödsorsaker i historiska data.

Aktiviteter under 2020

1. En referensgrupp med sakkunskap om historiska dödsorsaker utses, med uppdrag att bistå vid kodning av nya dödsorsaker.

2. Ta fram en dokumentation för användare som tillgängliggörs på swedpop.se

Tidsplan

Det grundläggande arbetet med harmonisering av dödsorsakskodning i SwedPop är avslutat. Fortsatt harmonisering av dödsorsaker sker löpande, när nya data läggs till, efter de principer som har tagits fram inom denna modul.

1b. Yrken

Nuläge

Under 2019 har arbetet med harmonisering av yrkeskodning i SwedPop slutförts. Ett gemensamt kodschema för yrkeskodning, *Yrkeskodning i SwedPop 1.0*, har tagits fram och fastställts av ledningsgruppen 2019-09-06. Rutiner för fastställande av nya yrkeskoder, när behov av sådana uppstår under det löpande arbetet, har tagits fram.

Aktiviteter under 2020

1. Implementering av fastställt kodschema för kodning av nya data. Utförs av varje databas för sig.
2. Ta fram en användardokumentation av de principer som används för yrkeskodning inom SwedPop, som tillgängliggörs på swedpop.se

Tidsplan

Fortsatt harmonisering av yrkeskodning sker löpande när nya data läggs till, enligt de principer som har fastställts av ledningsgruppen.

1c. Församlingskodning

Nuläge

Arbetet med harmonisering av församlingskodning har fortsatt under 2019 under ledning av en särskild arbetsgrupp. En gemensamt kodschema har tagits fram, som tar hänsyn till förändringar över tid, samt en översättningstabell från befintlig kodning. Kodschemat fastställs av ledningsgruppen 2019-10-29. Implementering och omkodning kommer därefter att ske av varje databas för sig. Harmoniseringen av församlingskodning är därmed slutförd.

Aktiviteter under 2020

1. Ta fram en dokumentation för användare som tillgängliggörs på swedpop.se

Tidsplan

1. Fortsatt harmonisering av församlingskodning sker löpande när nya data läggs till, enligt de principer som fastställs.

B. IDS-konvertering av databaserna:

Nuläge

Arbetet med IDS-konvertering av databaser i SwedPop, som startade redan under samordningsprojektet har fortsatt under 2018-2019. Arbetet har bestått av en omfattande probleminventering och analys med syfte att hitta gemensamma definitioner och principer för konverteringsarbetet till IDS 4.0. Varje organisation i konsortiet arbetar vidare med att utveckla egna verktyg för konvertering, då databaserna är så olika till sin karaktär.

Under 2019 har arbetsgruppen fokuserat på att fastställa detaljerade metadata för ett gemensamt databasformat. SEDD, CEDAR och SweGens har vardera konverterat en första version av sina data som har utvärderats och kommer att vidareutvecklas till en slutlig version under höstterminen 2019. Arbetet är något försenat i förhållande till tidigare upprättad tidsplan då det behöver samordnas med det pågående arbetet i modul 3, se nedan.

Aktiviteter under 2020

1. Konvertering av GOPP och Rotemansdatabasen till gemensamt dataformat
2. Framtagande av metadata och systemdokumentation för konvertering

Tidsplan

- Konvertering av DDB, SEDD, GOPP, SweGens och Rotemansdatabaserna till gemensamt dataformat ska vara genomförd under 2020

Modul 3

A. Utveckling av ett enkelt uttagsverktyg på webportalen

Nuläge

Arbetet startade redan under 2018 med att ta fram riktlinjer för vilken funktionalitet verktygets ska ha och har fortsatt under 2019 med framtagande av en detaljerad kravspecifikation. Utvecklingen av verktyget sker i Umeå och SEDD och Riksarkivet deltar i planeringen av det pågående arbetet. Utvecklingen av uttagsverktyget har tagit mer tid i anspråk än beräknat, då den har visat sig att den underliggande databasstrukturen behöver kompletteras. Hur detta bäst kan ske har diskuterats under året, vilket har försenat arbetet. En reviderad projektplan fastställs av ledningsgruppen 2019-10-29. Avsikten är att ett uttagsverktyg som kan användas för nedladdning av dataset för forskning ska finnas på plats under 2020. En referensgrupp av forskare har utsetts och kommer att knytas till det fortsatta utvecklingsarbetet.

Aktiviteter under 2020

1. Fortsatt utveckling av ett uttagsverktyg med full funktionalitet.

Tidsplan:

- Under 2020 ska ett uttagsverktyg med full funktionalitet finnas tillgängligt på SwedPop.se, där forskare kan hämta hem harmoniserade data från samtliga databaser (Delmål 2020:a)

B. Utveckling av webbportal

Nuläge

Under 2019 har swedpop.se har uppdaterats med aktuell information. I övrigt har arbetet med att utveckla webbsidan inte prioriterats, i väntan på uttagsverktyget.

Aktiviteter under 2020

1. En nyhetssida utvecklas och swedpop.se marknadsförs i olika kanaler.

Tidsplan:

- En aktuell och uppdaterad webbsida ska finnas tillgänglig under första kvartalet 2020.

Modul 4

Utveckling av verktyg, för att länka samman databaserna i SwedPop

Nuläge

Arbetet startade hösten 2018 med en workshop för att ta fram en kravspecifikation för det länkningsverktyg som ska utvecklas och en plan för det fortsatta arbetet. Under 2019 har arbetet påbörjats med att utveckla verktyg för länkning av folkräkningar till Sveriges Dödbok. Därefter ska verktyg och metoder utvecklas fram för att länka övriga databaser till Sveriges Dödbok, för att slutligen länkas till varandra. Det ursprungliga målet var att verktyget skulle vara klart under 2019, men metodologiskt motiverade omprioriteringar under arbetets gång har gjort att det har blivit något försenat i förhållande till tidigare plan. Ett utvärderingssteg har lagts in i processen vilket i korthet innebär att ett analyssteg har lagts in efter det första länkningssteget (folkräkningar-Sveriges Dödbok), för att se om eventuella justeringar behöver göras.

Aktiviteter under 2020

1. Utveckling av verktyg för att länka samman databaserna i SwedPop slutförs (Delmål 2020:b, kvarstår från 2019)

Tidsplan:

- Arbetet ska vara avslutat under 2020.

Modul 5

Utvidga GOPP-databasen med nya data

Nuläge

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. 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 pågår.

Aktiviteter under 2020

1. Registreringen ur personakterna avslutas och registreringen ur församlingsböckerna fortsätter.
2. Komplettering av databasen med information om inkomst och yrke för personer observerade 1943.

Tidsplan

Datainsamlingen kommer att pågå hela bidragsperioden 2018-2022 och kommer inte att hinna avslutas under den femårsperiod som har beviljats medel.

Modul 6

Komplettering och länkning av data i Rotemansdatabasen

Nuläge

När Rotemansdatabasen 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.

Aktiviteter under 2020

1. Fortsatt komplettering och länkning av tidigare utelämnade data i Rotemansdatabasen.

Tidsplan

Kompletteringen av data och länkning är genomförd 2021.

Modul 7

Komplettering av dödsorsaker i Rotemansdatabasen

Nuläge

För att Rotemansdatabasens information om dödsorsaker ska bli fullständig behöver en kompletterande inregistrering ske av uppgifter i Stadsläkarens arkiv. Under 2019 inscannades samtliga dödsorsaksintyg och länkning av tidigare olänkade poster påbörjades. En analys har också gjorts av var kompletteringar behöver ske. I nuläget har 150 000 av 220 000 dödsorsaksintyg transkriberats och omkring hälften av dessa har hunnit länkas till databasen.

Aktiviteter under 2020

1. Fortsatt registrering och länkning av tidigare ej registrerade dödsorsaksintyg. (Delmål 2020:c)

Tidsplan

Arbetet beräknas vara slutfört vid halvårsskiftet 2020.

Modul 8

Komplettera Rotemansdatabasen med ekonomiska data

Nuläge

För att öka den komparativa forskningspotentialen i SwedPop kommer Rotemansdatabasen att kompletteras med ekonomiska data. Arbetet kommer att inledas med validering och länkning av 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. Arbetet startade under hösten 2019, något försenat, då ett större fokus har lagts på aktiviteter inom modul 6 och 7 och kommer att skjuta över på 2020.

Aktiviteter under 2020

1. Fortsatt validering av registret för 1901 års huvudförteckning
2. Fortsatt verifiering av tidigare utförd länkning av 1901 års huvudförteckning

Tidsplan

Aktiviteterna ska vara genomförda under vid halvårsskiftet 2020.

Modul 9

Komplettering av SEDD med ekonomiska data för Landskrona 1905-1947

Nuläge

Motsvarande ekonomiska data från inkomst- och taxeringsregister för tidsperioden 1948-1968 finns sedan tidigare inkluderade i databasen och används för forskning. Arbetet inom denna modul avser en komplettering med data för tidsperioden 1905-1947, för att öka databasens forskningsvärde. Hittills har inkomstlängderna för åren 1915, 1920, 1925, 1930, 1935, 1938, 1940 och 1945-46 registrerats och länkats till databasen. Strategin i modulen har justerats 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 mellanliggande år.

Aktiviteter under 2020

1. Inkomst- och taxeringsuppgifter för åren 1906, 1910, 1942 och 1944 kommer att registreras och länkas till SEDD (Delmål 2020:d)
1. Yrkesuppgifter från längderna kodas.

Tidplan

Arbetet kommer att pågå under hela bidragsperioden 2018-2022.

Modul 10

Komplettering av SEDD med befolkningsdata för Landskrona 1895-1910

Nuläge

Befolkningsdata för Landskrona 1910-1968, och från omkringliggande församlingar 1813-1968 har inkluderats i SEDD inom ramen för tidigare projekt.

Aktiviteter under 2020

1. Data från församlingsböckerna för perioden 1904-1910 registreras och länkas till befintlig databas
2. SEDD-databasen har kompletterats med information om fars yrke för individer i Landskrona födda 1905-1947 (Delmål 2010:e)
3. Kontroll av registreringen mot födelse-, död-, giftermåls-, och flyttningslängder
4. Kodning av tillkomna yrken och dödsorsaker.

Tidplan

Arbete kommer att pågå under tidsperioden 2020-2022.

Modul 11

Komplettering av DDB med befolkningsdata för Nysätra församling

Nuläge

Sedan 2008 har befolkningsdatabasen POPLINK varit under uppbyggnad vid DDB i Umeå, inom ramen för tidigare projekt. Avsikten är att databasen på sikt ska omfatta befolkningen i Västerbotten 1700-1950. Att inkludera befolkningsdata från Nysätra är av strategisk betydelse för att knyta samman databasens två stora befolkningskluster, Skellefteå och Umeå. Arbetet startade under 2019 och ligger något före fastställd tidsplan

Aktiviteter under 2020

1. Fortsatt inregistrering av befolkningsdata från Nysätra församling

Tidplan

Inregistreringsarbetet startar 2019 och kommer att avslutas under 2020.

Modul 12

Komplettering av DDB med ekonomiska data för Lövångers församling

Nuläge

För att öka den komparativa forskningspotentialen i SwedPop kommer DDB:s befolkningsdatabaser att kompletteras med ekonomiska data. En förstudie påbörjades under hösten 2018. Arbetet har försenats av de aktuella längderna endast var tillgängliga i original och därmed inte kunde lånas ut. De kommer att scannas under hösten 2019. Därmed kan också förstudien avslutas och utvecklingen av inregistreringsformulär påbörjas. Inför framtiden bevakas också arbetet i ett pågående projekt vid universitetet i Odense, med automatisk inläsning av ekonomiska data.

Aktiviteter under 2020

1. Förstudie avslutas
2. Utveckling påbörjas av ett nytt formulär i inregistreringssystemet PERSONA för insamling av data från inkomst- och taxeringslängder (Delmål 2020:f)

Tidplan

Utvecklingen av ett nytt formulär sker under 2019-2020.

Datainsamlingen genomförs under 2022.

Modul 13

Inkludering av 1940 års folkräkning i SweCens-databasen

Nuläge

Idag finns folkräkningar från 1880, 1890, 1900 och 1910 tillgängliga för forskning i SweCens-databasen. Arbetet med 1930 års folkräkning, finansierat av andra projekt, är snart färdigt. Under 2018 utvecklades ett nytt formulär i inregistreringssystemet PERSONA avsett för folkräkningar som är driftssatt. Inscanningen av källmaterialet är genomförd och datainsamling har påbörjats. De första församlingarna är nu komplett inregistrerade. I samarbete med Sveriges Släktforskarförbund startade under hösten 2019 ett storskaligt citizen science-projekt där frivilliga släktforskare deltar i inregistreringsarbetet. Förhoppningen är att datainsamlingen därigenom ska kunna ske i snabbare takt än vad som tidigare planerats.

Aktiviteter under 2020

1. Fortsatt inregistrering av data i 1940 års folkräkning.
2. Rekrytering av ytterligare registrerare i Sveriges Släktforskarförbunds regi

Tidplan

Datainsamling påbörjas 2019 och kommer att pågå hela bidragsperioden, till och med 2022. Den kommer med nuvarande resurser inte att hinna avslutas under den tid projektet har finansiering.

Modul 14

Inkludering av 1950 års folkräkning i SwedPop

Nuläge

Folkräkningen från 1950 har transkriberats och digitaliserats för genealogiska ändamål av Arkiv Digital (AD). Enligt den ursprungliga projektplanen skulle dessa data genom ett samarbetsavtal mellan AD och Riksarkivet inkluderas i SweCens-databasen och tillgängliggöras för forskning. Innan projektstart sades samarbetsavtalet upp av AD och en ny överenskommelse med RA har inte kunnat nås. Under 2019 har ledningen för SwedPop-projektet på styrgruppens uppdrag förhandlat med AD angående möjligheten att mot ersättning få tillgång till folkräkningen för projektets räkning. Förhandlingarna har varit positiva och under hösten 2019 förbereds en upphandling av data som förhoppningsvis kan genomföras innan årsskiftet.

Aktiviteter under 2020

Upparbetning och kodning av 1950-års folkräkning samt anpassning till användning för forskningsändamål (Delmål 2020:g)

Tidplan

Arbetet ska vara avslutat under 2020.

Modul 15

Sveriges dödbok

Nuläge

Genom ett avtal med Sveriges Släktforskarförbund kommer SwedPop att för forskningsändamål få tillgång till en kopia av den version av databasen Sveriges Dödbok, som publiceras i november 2018. Sveriges dödbok innehåller information om alla individer som har avlidit i Sverige från 1860 och fram till idag. Databasen kommer att harmoniseras, upparbetas och länkas samman med övriga databaser i SwedPop. Den kommer också att konverteras till IDS-format. Komplettering och harmonisering pågår inom ramen för andra projekt. Upparbetning och validering har startat i mindre skala under 2018 och 2019.

Aktiviteter under 2020

1. Kodning av geografisk information
2. Länkning till folkräkningar och övriga databaser i SwedPop påbörjas.

Tidplan

Arbetet kommer att pågå under tidsperioden 2020-2022