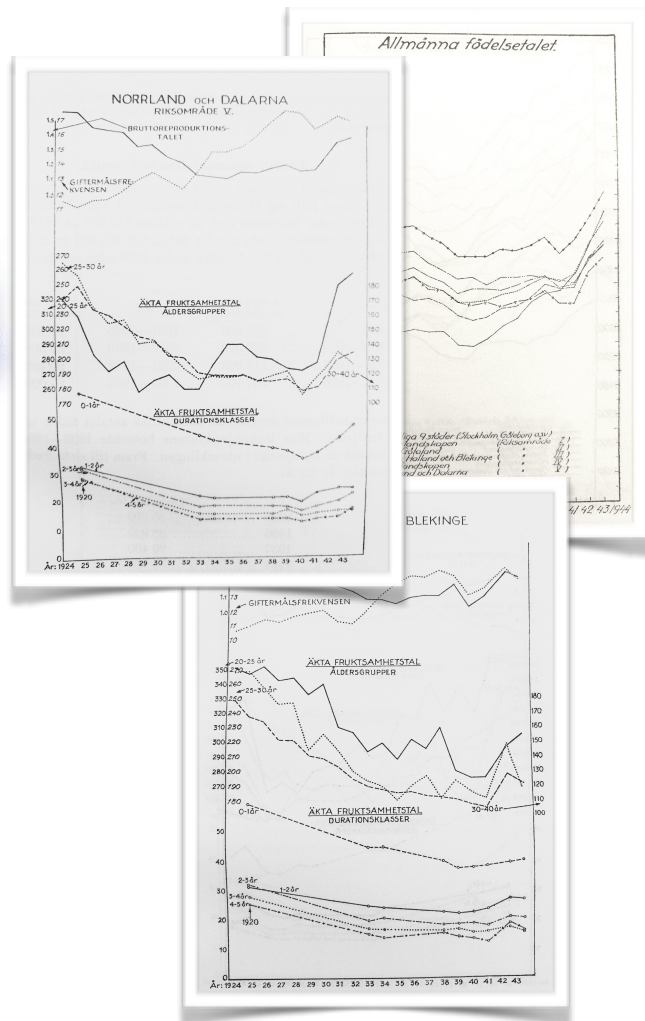


SwedPop

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

Sveriges befolkningsdata saknar motstycke i världen. Kontinuerliga data över befolkningen, med social och ekonomisk information, finns tillgängliga från 1600-talet fram till idag. Dessa data ligger till grund för flera stora forskningsdatabaser som var för sig är värdefulla resurser för forskning inom många vetenskapsområden. Trots den rika datatillgången har den stora forskningspotentialen hos svenska befolkningsdata långt ifrån utnyttjats till fullo, delvis beroende på bristande samordning och harmonisering av data samt skillnader i uppbyggnad och struktur mellan olika databaser. I den nya forskningsinfrastrukturen SwedPop kommer de viktigaste historiska befolkningsdatabaserna i Sverige att samordnas till en gemensam resurs, med fokus på perioden från 1700-talet till 1960-talet. Denna samordning kommer att väsentligt öka, förbättra och förenkla forskares tillgång till högkvalitativa svenska befolkningsdata för forskning och möjliggöra longitudinella studier på individnivå över flera hundra år.

SwedPop kommer att tillgängliggöra demografiska, sociala och ekonomiska data på individnivå för stora delar av Sveriges 1800- och 1900-talsbefolkning. Detaljerade data på individ och familjenivå över hela livsförloppet och i många generationer ger unika förutsättningar för att studera hur människors liv, hälsa, sociala och ekonomiska levnadsvillkor formats och förändrats i samspel med samhällets utveckling, från ett traditionellt bondesamhälle fram till dagens postindustriella välfärdssamhälle. Eftersom SwedPop omfattar data från skilda delar av Sverige, både stad och landsbygd samt olika socioekonomiska miljöer, ökar också möjligheterna att göra jämförande studier, till exempel av regionala skillnader. Inom det medicinska området möjliggör den nya infrastrukturen fördjupade analyser av dödlighetsnedgången under 1800- och 1900-talen. Genom länkning till andra forskningsregister möjliggörs studier av ett stort antal andra aktuella medicinska forskningsfrågor.

Kärnan i SwedPop utgörs av fem stora forskningsdatabaser: Kyrkboksdata baserna 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. Den nationella SweCens-databasen baseras på de svenska historiska folkräkningarna och GOPP är en panelstudie baserad på folkbokföringsdata från Göteborg. Tillsammans kommer de att erbjuda sammanhängande information om den svenska befolkningen med regionala nedslag, från tidigt 1800-tal och fram till 1950. I förlängningen kommer även andra databaser att kunna knytas till SwedPop, med syfte 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.

Flera av de ingående databaserna 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 att fokusera på följande områden:

- Samordning av data i form av kodning av yrken, dödsorsaker och geografiska enheter
- Strukturering av data enligt gemensam databasstruktur (IDS) med gemensam 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

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 2018-xx-xx 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-xx.

Under 2019 är målsättningen att arbetet ska leda fram till följande delresultat:

			Modul
	a	Kodningen av dödsorsaker, yrkesuppgifter och församlingsinformation har harmoniserats i de ingående databaserna.	2
	b	SweCens databasen har konverterats till IDS-struktur	2
	c	Ett länkningsverktyg har utvecklats, som är utformat för att länka samman de ingående databaserna.	4
	d	GOPP-databasen har kompletterats med information om utvalda personer observerade 1943 och deras hushåll.	5
	e	SEDD-databasen har kompletterats med inkomstdata för åren 1925, 1930 and 1935.	9
	f	Inregistreringssystemet PERSONA har anpassats till arbete med digitalisering av 1940 års folkräkning.	13
	g	1950 års folkräkning har kodats och länkats till övriga folkräkningar, samt till övriga databaser i SwedPop.	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 2019.

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. Dödsorsaker i Umeå och Lund är kodade enligt gemensamma principer och arbetet med harmonisering av dödsorsaker i Rotemansdatabasen har påbörjats. SweCens-databasen innehåller inga dödsorsaker. I GOPP-databasen sker huvuddelen av harmoniseringen i och med inkludering och kodning av nya data. Ingen särskild arbetsgrupp behövs för detta arbete.

Aktiviteter under 2019

1. Fortsatt kodning av dödsorsaker i Rotemansdatabasen (Delmål 2019:a)
2. Fortsatt kodning av dödsorsaker i GOPP-databasen (Delmål 2019:a)

Tidsplan

Dödsorsaker i samtliga ingående databaser ska vara kodade under 2019.

Fortsatt harmonisering av dödsorsaker sker löpande, när nya data läggs till, efter de principer som har tagits fram inom denna modul.

1b. Yrken*Nuläge*

Den arbetsgrupp som tillsattes under samordningsprojektet har fortsatt sitt arbete under 2018. En dokumentation av den yrkeskodning som har tillämpats inom varje databas har tagits fram och en probleminventering har gjorts. En workshop genomförs under hösten 2018 med syfte att fastställa principer för standardisering och en gemensam tillämpning för kodning av yrken enligt HISCO.

Aktiviteter under 2019

1. Implementering av fastställt kodschema för SwedPop-data (Delmål 2019:a)
2. Fortsatt kodning av yrken enligt fastställda principer. Utförs av varje databas för sig.
3. Fastställa rutiner för hur vi på längre sikt ska hantera skillnader i yrkeskodning, som kan uppkomma när nya data inregistreras.

Tidsplan

Fortsatt harmonisering av yrkeskodning sker löpande, när nya data läggs till, efter de principer som har tagits fram inom denna modul.

1c. Församlingskodning*Nuläge*

Arbetet med harmonisering av församlingskodning har blivit något försenat, men kommer att starta under senhösten 2018 med en gemensam workshop. En arbetsgrupp finns tillsatt för att leda arbetet med att ta fram en gemensam kodning över tid, samt en översättningstabell från befintlig kodning. Omkodning kommer därefter att ske av varje databas för sig. Arbetsgruppens uppdrag sträcker sig över perioden 2018-2019.

Aktiviteter under 2019

1. Probleminventering och analys med utgångspunkt i listor på befintlig kodning av församlingar som har tagits fram av respektive databas
2. Ta fram gemensamma kodscheman för olika tidsperioder (steg 1) (Delmål 2019:a)
3. Ta fram en översättningstabell för geografisk kodning över tid (steg 2) (Delmål 2019:a)

Tidsplan

- Probleminventering och analys ska vara klar vid årsskiftet 2018/2019.
- Arbetet med att ta fram kodscheman och översättningstabell ska vara klart under 2019.
- Fortsatt harmonisering av församlingskodning sker löpande, när nya data läggs till, efter de principer som har tagits fram inom denna modul.

B. IDS-konvertering av databaserna:

Nuläge

Arbetet med IDS-konvertering av databaser i SwedPop har fortsatt under 2018, under ledning av en projektgrupp. Gruppens arbete har hittills inriktats på probleminventering och analys med syfte att hitta gemensamma definitioner och principer för konverteringsarbetet till IDS 4.0. Varje organisation i konsortiet arbetar med att utveckla egna verktyg för konvertering, då databaserna är så olika till sin karaktär. En workshop anordnades i maj 2018 med medverkan av IDS-experterna George Alter (USA) och Kees Mandemakers (Nederländerna). Under hösten 2018 fattas beslut om vilka principer som ska gälla för ett första tillgängliggörande av data på swedpop.se vid årsskiftet 2018/2019. I nuläget avvaktar vid med IDS-extension-tabeller utan fokuserar på standardtabeller. Vi ser också att det fortsatta arbetet med IDS-konvertering behöver ske i nära samarbete med utvecklingen av uttagsverktyg i modul 3, vilket kan komma att innebära att modifieringar kan behöva göras under projektets gång.

Aktiviteter under 2019

1. Utveckling av databasspecifika verktyg för konvertering av GOPP- SweCens- och Rotemansdatabaserna (Delmål 2019:b)
2. Fortsatt arbete med IDS-konvertering av SweCens utifrån de principer som fastställs hösten 2018 (Delmål 2019:b)
3. Systemdokumentation för omkodningen till IDS tas fram i samband med att IDS-konverteringen av databaserna färdigställs.

Tidsplan

- Konvertering av DDB, SEDD, SweCens och Rotemansdatabaserna till IDS 4.0 ska vara utförd under 2019.

Modul 3

A. Utveckling av ett enkelt uttagsverktyg på webportalen

Nuläge

Arbetet startade under 2018 med att ta fram riktlinjer för vilken funktionalitet uttagsverktyget ska ha i en första version och lägga upp en plan för det fortsatta arbetet. Utvecklingen av verktyget kommer att ske i Umeå och SEDD och Riksarkivet kommer att finnas med som kravställare och testare. En första kravspecifikation har upprättats och utvecklingsarbetet kommer att starta under senhösten, något försenat i relation till tidigare upprättad tidsplan.

Aktiviteter under 2019

1. Upprätta en kravspecifikation för den funktionalitet uttagsverktyget ska ha i sin slutversion och ta fram en plan för den fortsatta utvecklingen.
2. 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 data från samtliga databaser.

B. Utveckling av webbportal*Nuläge*

Under 2018 har ledningsgruppen arbetat vidare med struktur och innehåll och swedpop.se har uppdaterats med aktuell information.

Aktiviteter under 2019

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

Tidsplan:

- En aktuell och uppdaterad webbsida ska finnas tillgänglig under första kvartalet 2019

Modul 4

Utveckling av verktyg, för att länka samman databaserna i SwedPop*Nuläge*

Arbetet startade hösten 2018 med en workshop med syfte att ta fram en kravspecifikation för det länkningsverktyg som ska utvecklas och en plan för arbetet.

Aktiviteter under 2019

1. Utveckling av länkningsverktyg enligt framtagna kravspecifikation (Delmål 2019:c)

Tidsplan:

- Ett länkningsverktyg för att länka samman de ingående databaserna har utvecklats under 2019.

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. Inregistreringen av nya data kommer att göras av Riksarkivet i Göteborg, som ett arvodesuppdrag.

Aktiviteter under 2019

1. Komplettering av databasen med information om utvalda personer observerade 1943 och deras hushåll (Delmål 2019:5)

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 har genomförts under 2018. Kompletteringsarbetet kommer att starta under 2019.

Aktiviteter under 2019

1. Komplettering och länkning av data i Rotemansdatabasen, enligt tidigare utförd validering.

Tidsplan

Kompletteringen av data och länkning är i huvudsak genomförd under 2019.

Modul 7

Komplettering av dödsorsaker i Rotemansdatabasen

Nuläge

För att komplettera Rotemansdatabasen med dödsorsaker behöver en kompletterande inregistrering ske av uppgifter i Stadsläkarens arkiv. I nuläget har 150 000 av 220 000 dödsorsaksintyg transkriberats och omkring hälften av dessa har hunnit länkas till databasen. Inskanning av återstående dödsorsaksintyg har upphandlats och utförts under 2018

Aktiviteter under 2019

1. Upparbetning av tidigare inregistrerade data
2. Fortsatt registrering av återstående dödsorsaker

Tidsplan

Alla förberedelser för inregistrering och länkning av nya data är färdiga under 2018 och den kompletterande inregistreringen har påbörjats. Arbetet beräknas kunna slutföras under 2020.

Modul 8

Komplettera Rotemansdatabasen med ekonomiska data

Nuläge

??

Aktiviteter under 2019

???

Tidsplan

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.

Aktiviteter under 2019

1. Inkomst- och taxeringsuppgifter för 1925, 1930 och 1935 är inregistrerade, länkade och inkluderade i SEDD (Delmål 2019:e)

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. Under 2018-2019 pågår datainsamling för tidsperioden 1910-1930 inom ett projekt finansierat av Riksbankens jubileumsfond.

Aktiviteter under 2019

1. Inga aktiviteter är planerade under 2019.

Tidplan

Arbetet kommer att starta först 2020.

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.

Aktiviteter under 2019

1. Påbörjad inregistrering av befolkningsdata från Nysätra församling

Tidplan

Inregistreringsarbetet startar 2019 och kommer att pågå till och med 2021.

Modul 12

Komplettering av DDB med ekonomiska data för Lövångers församling

Nuläge

Inga ekonomiska data finns ännu inkluderade i DDB:s befolkningsdatabaser. Avsikten med denna komplettering är att öka den komparativa forskningspotentialen i SwedPop. En förstudie genomfördes under hösten 2018.

Aktiviteter under 2019

1. Utveckling påbörjas av ett nytt formulär i inregistreringssystemet PERSONA för insamling av data från inkomst- och taxeringslängder

Tidplan

Utvecklingen av ett nytt formulär sker under 2019-2020.

Datainsamlingen sker 2022.

Modul 13

Inkludering av 1940 års folkräkning i SweCens-databasen

Nuläge

Idag finns folkräkningar från 1880, 1890 och 1900 tillgängliga för forskning i SweCens-databasen. Arbete pågår med 1930 års folkräkning, finansierat av andra projekt. Inscanning av 1940 års folkräkning pågår. Utveckling av ett nytt formulär i inregistreringssystemet PERSONA avsett för folkräkningar startade under 2018.

Aktiviteter under 2019

1. Inregistreringssystemet PERSONA har anpassats till arbete med 1940 års folkräkning (Delmål 2019:f)
2. Inregistrering av data i 1940 års folkräkning påbörjas

Tidplan

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

Modul 14

Inkludering av 1950 års folkräkning i SweCens-databasen

Nuläge

Folkräkningen från 1950 har transkriberats och digitaliserats för genealogiska ändamål av Arkiv Digital. Tyvärr har problem uppstått med tidigare tecknat samarbetsavtal med Arkiv Digital. Riksarkivet har ännu inte fått tillgång till den tidigare utlovade kopian för forskningsändamål, som i upparbetad form var avsedd att inkluderas i SweCens-databasen och länkas till övriga folkräkningar. Fortsatta förhandlingar pågår.

Aktiviteter under 2019

Detta delmål kommer inte att kunna genomföras enligt tidigare upprättad tidsplan (Delmål 2019:g)

Tidplan

I dagsläget oklart när arbetet kan påbörjas.

Modul 15

Sveriges dödbok

Nuläge

Genom ett avtal med Sveriges Släktforskarförbund kommer Riksarkivet 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 upparbetas och länkas samman med övriga databaser i SwedPop.

Aktiviteter under 2019

Inga aktiviteter är planerade under 2019.

Tidplan

Arbete inom denna modul kommer att starta först 2020.