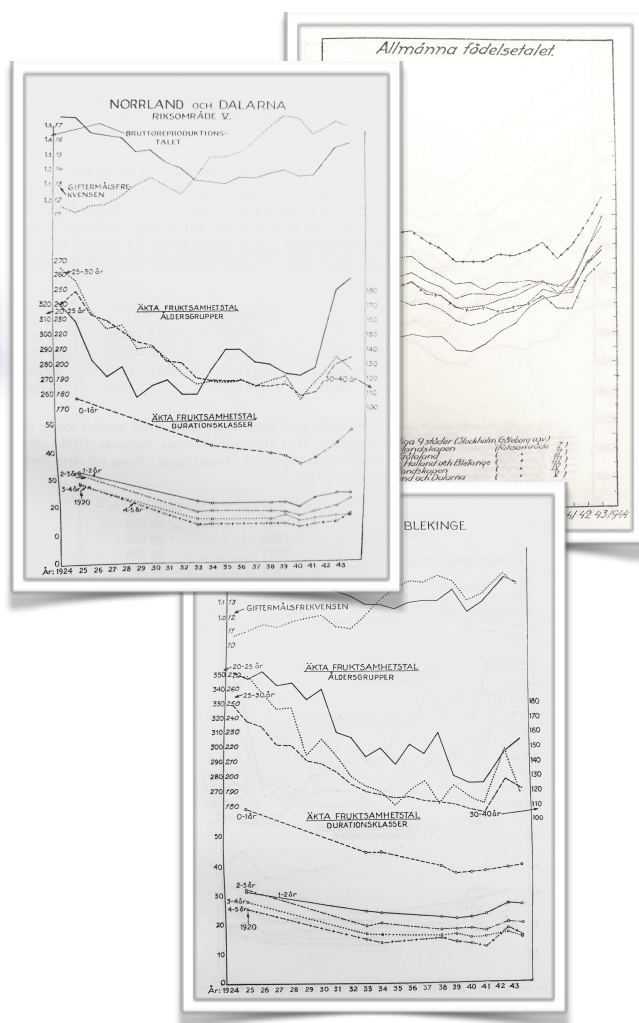


SwedPop

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 befolkningen i 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 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 I skånska församlingar 1680-1968.
- **Riksarkivet: SweCens-databasen** Individdata från svenska folkräkningar 1880-1950. Tvärsnitt vart tionde år för hela riket.
- **Stockholms stadsarkiv: Rotemansarkivet** Livsförloppsdata för Stockholms befolkning 1877-1926.

SwedPop: Svenska befolkningsdatabaser för forskning 2018-2022

1. Inledning

Bakgrund

SwedPop beviljades 2017 bidrag från Vetenskapsrådet inom ramen för utlysningen *Bidrag till infrastruktur av nationellt intresse 2017*. 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 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.

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). Både delmål och resultat uppnådda under 2018 kommer att diskuteras under respektive modul.

2. Verksamheten under 2018

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 Rolf Källman	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 2018 sammanträtt vid två tillfällen och ledningsgruppen har haft sju protokollförda sammanträden.

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

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)

Med några få undantag har arbetet följt upprättad verksamhetsplan och uppsatta budgetramar. Följande förändringar har gjorts av den ursprungliga arbetsplanen:

- I modul 2 har tanken på ett gemensamt konverteringsverktyg fått överges, då det bedömdes rimligare att arbeta vidare med databasspecifika verktyg för konvertering.
- I modul 3 har tidsplan och målbild behövt revideras (se vidare under modul 3). Ambitionen att lägga upp ett enkelt uttagsverktyg vid årsskiftet 2018/19 fick överges, då harmoniseringen ännu inte var helt avslutad och det bedömdes bättre att vänta med tillgängliggörandet till dess verktyget har fått en önskad funktionalitet.
- I modul 13 har vissa arbetsmoment tidigare lagts. Av denna anledning har medelförbrukningen överstigit årets budget, men detta kommer att justeras under 2019. Sammantaget utgör dessa avsteg från verksamhetsplanen smärre kursändringar som inte kommer att få några omfattande konsekvenser för projektet i ett längre perspektiv. De kommer inte heller att ha någon större budgetpåverkan.

Ett mer bekymmersamt avsteg från verksamhetsplanen är att arbetet i modul 14 ännu inte har kunnat starta. Till följd av oförutsedda händelser utanför vår kontroll har vi ännu inte fått tillgång till 1950 års folkräkning (se vidare under modul 14). Frågan har diskuterats i styrgruppen och det finns en plan för hur vi ska gå vidare, men färdigställandet av arbetet i modul 14 kommer att bli försenat.

I några moduler har avsatta medel inte kunnat förbrukats som planerat enligt budget, till följd av att aktiviteter av olika skäl inte har kunnat genomföras. Det rör bland annat modulerna 2 och 14 som har diskuterats ovan. I modulerna 6 och 7 har vissa planerade aktiviteter fått skjutas framåt i tiden, då personalresurser har varit uppbundna av Stockholms stadsarkivs flytt till nya lokaler. Även kodningen av dödsorsaker i modul 2 har försenats något av flytten.

Kommunikation och information

Webbportalen swedpop.se har utvecklats under året och innehåller nu fördjupad information om projektet och om parterna i konsortiet. Utvecklingsarbetet kommer att fortsätta under projekttiden. Den 15 mars hölls ett seminarium i samarbete med Riksarkivet

där SwedPop presenterades för intresserade forskare och företrädare för arkivsektorn. Infrastrukturen presenterades även vid en session vid Social Science History Association Annual Meeting i Phoenix, USA den 8-11 november, tillsammans med liknande pågående initiativ i Europa och USA. Etablerade kontakter och samarbeten finns med liknande befolkningsdatabaser i övriga nordiska länder samt internationellt.

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 finns uppsatta för denna modul. Verksamheten följer upprättad budget.

Modul 2

A. Harmonisering av kodning:

1. Harmonisering av variabler

- Dödsorsaker
- Yrke
- Församlingskodning

1a: Dödsorsaker

Harmoniseringen av dödsorsakskodningen i databaserna till ICD10-standard påbörjades under samordningsprojektet 2016-2017 och fortsatte under 2018. Gemensamma principer för kodning har utarbetats och implementerats. Vid årsskiftet var kodningen av dödsorsaker i DDB och SEDD avslutad. Förberedelser hade startat för kodningen av dödsorsaker i Rotemansdatabasen, dock något försenad i förhållande till tidigare upprättad tidsplan till följd av Stockholms Stadsarkivs flytt till nya lokaler. SweCens- och GOPP-databaserna innehåller inga dödsorsaker. Under året har SwedPop, via de medarbetare som arbetat inom denna modul, också tagit aktiv del i det utvecklingsarbete som pågår inom det europeiska samarbetsprojektet SHiP. Syftet är att ta fram en enhetlig europeisk standard för kodning av historiska dödsorsaker, ICD10H baserad på ICD10.

1b. Yrken

Arbetet med harmonisering av databasernas yrkeskodning har fortsatt under 2018 under ledning av den arbetsgrupp som tillsattes under samordningsprojektet 2016-2017. En dokumentation av den yrkeskodning som har tillämpats inom respektive databas har tagits fram och en probleminventering har gjorts. En workshop genomfördes under hösten 2018 med syfte att fastställa principer för standardisering och gemensam tillämpning av internationell standard för yrkeskodningen (HISCO). Ett kodschema för yrkeskodning inom SwedPop har utarbetats, som kommer att fastställas av ledningsgruppen under våren 2019.

1c. Församlingskodning

Arbetet med harmonisering av församlingskodning startade något försenat under senhösten 2018 med en workshop för probleminventering och analys, under ledning av en tillsatt arbetsgrupp. Arbetsgruppens uppdrag sträcker sig över perioden 2018-2019. Även om arbetet startade senare än planerat, är bedömningen att vi kommer att hålla upprättad tidsplan. Det betyder att ett fastställt kodschema med översättningstabell ska finnas tillgängligt under 2019. Omkodning kommer därefter att ske av varje databas för sig.

B. IDS-konvertering av databaserna:

Arbetet med IDS-konvertering av databaserna i SwedPop fortsatte under 2018, under ledning av tidigare tillsatt projektgrupp. Gruppens arbete har under året varit inriktat på probleminventering och analys, med syfte att fastställa gemensamma definitioner och principer för konverteringsarbetet till IDS 4.0. En workshop anordnades i maj 2018 med medverkan av IDS-experterna George Alter (USA) och Kees Mandemakers (Nederländerna). Principer för IDS-konvertering fastställdes av ledningsgruppen i oktober 2018.

Ursprungligen var avsikten att ta fram ett verktyg för IDS-konverteringen av databaserna, men detta har visat sig vara svårt att genomföra, till följd av att databaserna är så olika till sin karaktär. Istället kommer varje organisation i konsortiet att arbeta vidare med att utveckla egna databasspecifika verktyg för konvertering till IDS. Detta innebär att vi inte har klarat av att leverera ett av de resultat som hade utlovats under året: att ett verktyg för konvertering av databaserna till IDS-format skulle finnas tillgängligt vid slutet av 2018.

Bedömningen är dock att detta omtag inte kommer att leda till några förseningar av målet att samtliga databaser ska vara konverterade till IDS 4.0 under 2019. Data från DDB och SEDD finns i nuläget tillgängligt i IDS-format.

Modul 3

A. Utveckling av ett enkelt uttagsverktyg på webportalen

Arbetet startade under 2018 med att ta fram en kravspecifikation för uttagsverktyget och en plan för det fortsatta utvecklingsarbetet, som kommer att ske i Umeå. SEDD och Riksarkivet finns med som kravställare och testare och en referensgrupp av forskare kommer att knytas till arbetet. Utvecklingsarbetet startade under senhösten, något försenat

i relation till den ursprungliga tidsplanen då strategiska resurser var uppbundna i andra projekt. Målsättningen var att ett första enkelt uttagsverktyg skulle finnas tillgängligt på swedpop.se vid årsskiftet 2018-2019, för uttag av data från SEDD och DDB. Under hösten stod det dock klart att det skulle bli mycket svårt att ha ett användbart verktyg på plats inom utsatt tid. Arbetet med harmonisering av kodning var inte helt avslutat och tiden blev för kort för att hinna utveckla en funktionell produkt. Under hösten 2018 beslutade därför ledningsgruppen att avvakta med tillgängliggörandet av uttagsverktyget, till dess harmoniseringsarbetet är klart och till dess verktyget har en sådan funktionalitet att det fullt ut kan användas för nedladdning av forskningsbara dataset. Att lägga upp ett verktyg med begränsad funktionalitet bedöms vare sig gynna projektet eller de forskare som är potentiella användare av infrastrukturen. Resultatmålet att ett enkelt uttagsverktyg ska finnas tillgängligt på swedpop.se kommer alltså att förändras något i förhållande till tidigare projektplan.

B. Utveckling av webbportal

Under 2018 har webbplatsen swedpop.se vidareutvecklats. Den innehåller nu utförligare information om projektet och om de parter som finns i konsortiet. Arbetet har skett i samarbete med Inhouse-byrån vid Umeå universitet. Arbetet med webben kommer att fortsätta under 2019.

Modul 4

Utveckling av verktyg, för att länka samman databaserna i SwedPop

Arbetet startade hösten 2018 med en workshop med syfte att ta fram en kravspecifikation för det länkningsverktyg som ska utvecklas, samt att upprätta en plan för det fortsatta arbetet. Länkningen kommer att ta sin början under 2019.

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. Under 2018 har arbetet startat med att utvidga databasen med inregistrering av nya data. Personer som kan observeras 1943 följs fram till 1967 och nya urvalspersoner födda efter 1943 tillförs och följs till 1967. Inregistreringen av nya data görs av Riksarkivet i Göteborg, som ett arvodesuppdrag. Arbetet har under hösten försenats något till följd av sjukfrånvaro, men inte i en sådan omfattning att den övergripande tidsplanen kommer att påverkas.

Modul 6

Komplettering och länkning av data i Rotemansdatabasen

Förstudie, validering och dokumentation har påbörjats under 2018. Kompletteringsarbetet kommer att starta under 2019. Arbetet är något försenat i förhållande till tidigare upprättad tidsplan till följd av Stockholms stadsarkivs flytt till nya lokaler. Avsatta medel har inte heller fullt ut förbrukats under året.

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. Delar av arbetet inom denna modul har också blivit något försenat till följd av Stadsarkivets flytt och kommer att utföras under 2019. Under 2018 har de dödsorsaksintyg som ska transkriberas scannats in. Inregistrering och kodning kommer att fortsätta under kommande år. Detta innebär att avsatta medel inte har förbrukats enligt budget.

Modul 8

Komplettera Rotemansdatabasen med ekonomiska data

Arbetet startar först 2019.

Modul 9

Komplettering av SEDD med ekonomiska data för Landskrona 1905-1947

Motsvarande ekonomiska data från inkomst- och taxeringsregister för tidsperioden 1948-1968 finns sedan tidigare inkluderade i databasen och används för forskning. Arbetet under året har följt tidigare upprättad tidsplan och fastlagd budget, vilket innebär att inkomst- och taxeringsuppgifter för 1940, 1945 och 1947 nu är inregistrerade, länkade och inkluderade i SEDD.

Modul 10

Komplettering av SEDD med befolkningsdata för Landskrona 1895-1910

Arbetet startar först 2020.

Modul 11

Komplettering av DDB med befolkningsdata för Nysätra församling

Arbetet med att inkludera befolkningsdata från Nysätra startade under 2018 och följer upprättad tidplan och fastlagd budget.

Modul 12

Komplettering av DDB med ekonomiska data för Lövsångers församling

För att öka den komparativa forskningspotentialen i SwedPop kommer DDB:s befolkningsdatabaser att kompletteras med ekonomiska data. Arbetet startade med en förstudie under hösten 2018 och kommer att fortsätta under 2019.

Modul 13

Inkludering av 1940 års folkräkning i SweCens-databasen

Arbetet startade under 2018. Vissa arbetsmoment har tidigare lagts, vilket gör att arbetet ligger något före fastställd projektplan. Under året har fokus legat på förberedande arbetsmoment, varav den mesta tiden har ägnats åt att utveckla ett nytt inregistreringsformulär till mjukvaran PERSONA, som ska användas vid datainsamlingen. Formuläret har testats och är klart att ta i bruk för inregistrering av folkräkningar. Manualer för inregistrering har tagits fram och även övriga förberedelser är avslutade. Datainsamling startar 2019. Tidigareläggningen av systemutvecklingsinsatser har gjort att medelsförbrukningen har överskridit årets budget för denna modul, men detta kommer att regleras under 2019.

Modul 14

Inkludering av 1950 års folkräkning i SweCens-databasen

Folkräkningen från 1950 har transkriberats och digitaliserats för genealogiska ändamål av Arkiv Digital AB. Avsikten var att dessa data efter upparbetning skulle inkluderas i SweCens-databasen och tillgängliggöras för forskning. Detta garanterades när ansökan skrevs av ett undertecknat samarbetsavtal, som skulle ge Riksarkivet och därmed SwedPop tillgång till dessa data. Därefter har samarbetsavtalet, av anledningar som ligger utanför projektets kontroll, sagts upp av Arkiv Digital. Trots fortsatta förhandlingar har en ny överenskommelse inte kunnat nås. Av denna anledning har arbetet inom denna modul inte kunnat starta. Styrgruppen för SwedPop beslutade i november 2018 att konsortiet under våren 2019 ska göra en framställning till Arkiv Digital om utlämning av data. Andra möjligheter att få tillgång till befolkningsdata från 1950 kommer också att undersökas. Sammantaget innebär detta att arbetet inom denna modul inte har kunnat genomföras enligt tidigare upprättad tidsplan. Avsatta budgetresurser har därmed inte heller förbrukats. I dagsläget är det oklart när arbetet kan påbörjas.

Modul 15

Sveriges dödbok

Arbetet startar först 2020.

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 harmoniseringsarbetet ännu inte är helt avslutat och ett funktionellt uttagsverktyg ännu inte är på plats. Däremot har data från de databaser som ingår i SwedPop tillgängliggjorts för forskningsändamål under året. Data från SweCens-databasen ingår också i det internationella North Atlantic Population Project, där forskare fritt kan ladda ned dataset för forskning (<https://www.nappdata.org/napp/>). En översikt över antalet levererade dataset, antalet projekt och användare finns i den särskilda redovisningen av nyckeltal.

Under 2018 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 öppet 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 sökdatabas, men där Sveriges Släktforskarförbund har godkänt och förmedlat mer avancerade forskningsuttag.