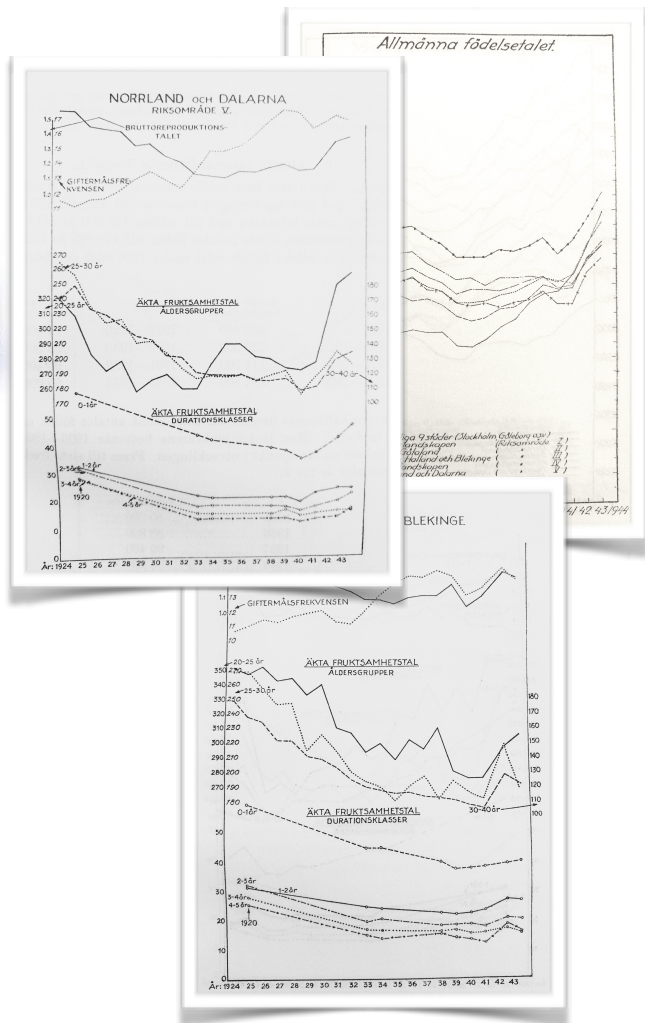


# SwedPop

## 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 detta 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. Möjligheten att länka data till moderna befolkningsregister skapar förutsättningar för longitudinella studier på individnivå under 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 skapar 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. Möjligheten att länka data till andra forskningsregister ger också förutsättningar för studier av en mängd 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 väsentligt öka forskningspotentialen för samtliga databaser och för infrastrukturen som helhet kommer utvecklingen av SwedPop att fokusera på följande områden:

- Samordna data i form av kodning av yrken, dödsorsaker och geografiska enheter
- Strukturera data enligt gemensam databasstruktur (IDS) med gemensam nomenklatur och kodningssystem
- Göra data tillgängliga för forskare från gemensam portal
- Länka individerna mellan de olika databaserna
- Registrera kompletterande data för att fylla luckor i tid och rum samt öka jämförbarheten mellan databaserna genom att införliva ekonomiska individdata
- Etablera SwedPop som nationellt centrum för äldre (före SCB-registren) befolkningsdata 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 vara Vetenskapsrådet tillhanda senast den 30 april 2018. Därefter ska verksamhetsplanen årligen vara Vetenskapsrådet tillhanda senast den 15 november.

Verksamhetsplanen för 2018 har utarbetats av den interimistiska ledningsgruppen för SwedPop, bestående av följande personer:

Elisabeth Engberg	CEDAR Umeå universitet
Maria Larsson	CEDAR Umeå universitet
Christer Lundh	Avdelningen för ekonomisk historia, Göteborgs universitet
Martin Dribe	Centrum för ekonomisk demografi, Lunds universitet
Mats Berggren	Riksarkivet
Anders Nordström	Riksarkivet

Johan Gidlöf                      Stockholms stadsarkiv

Gabriel Marawgeh              Stockholms stadsarkiv

Verksamhetsplanen är strukturerad utifrån de moduler som finns angivna i budget och arbetsplan för projektet.

## **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.) Dessa finns beskrivna i den övergripande arbetsplanen.

Under 2018 är målsättningen att arbetet ska leda fram till följande delresultat:

- a. Ett verktyg för konvertering av GOPP- SweCens- och Rotemansdatabaserna till IDS-format ska finnas tillgängligt vid slutet av 2018 (Modul 2)
- b. Ett första uttagsverktyg ska finnas tillgängligt på webportalen vid slutet av 2018 (Modul 3)
- c. Inkomst- och taxeringsuppgifter för 1940, 1945 och 1947 är inregistrerade, länkade och inkluderade i SEDD (Modul 9)

## 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 2018.

### Modul 2

---

#### A. Harmonisering av kodning:

##### 1. Harmonisering av variabler

- Dödsorsaker
- Yrke
- Församlingskodning

##### 1a: Dödsorsaker

###### *Nuläge*

Kodningen av dödsorsaker har påbörjats under samordningsprojektet 2016-2017 och fortsätter under 2018. Harmonisering sker i och med kodning, då ingen omfattande kodning av dödsorsaker tidigare har gjorts i de olika databaserna. Ingen särskild arbetsgrupp behöver tillsättas.

###### *Aktiviteter under 2018*

1. Fortsatt kodning av dödsorsaker i Umeå och Lund
2. Stockholms stadsarkiv skickar dödsorsaker till Umeå
3. Fortsatt kodning av dödsorsaker hos Stockholms stadsarkiv
4. Ett webinarium hålls med en presentation av de principer för kodning av dödsorsaker som tillämpas vid DDB i Umeå och en dokumentation sänds ut till alla parter.

###### *Tidsplan*

Dödsorsaker i Umeå och Lund ska vara kodade under 2018.

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 fortsätter sitt arbete, med Annika Westberg som sammankallande. En workshop genomfördes under hösten och uppföljande arbete pågår hos samtliga parter.

#### *Aktiviteter under 2018*

1. Ny fil skickas med yrkessträngar och kod skickas till Annika Westberg senast den 30 januari.
2. Arbetet med fortsatt dokumentation av den yrkeskodning som idag tillämpas i respektive databas.
3. Ta fram principer för standardisering och en gemensam tillämpning för kodning av yrken enligt HISCO.
4. Påbörja kodning av yrken enligt fastställt kodschema. Utförs av varje databas för sig.

#### *Tidsplan*

Kodschema för yrkeskodning klart under 2018.

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 ännu inte startat. En arbetsgrupp med uppdrag att hitta en gemensam kodning över tid, samt att ta fram en översättningstabell tillsätts. Omkodning kommer därefter att ske av varje databas för sig. Mats Berggren är sammankallande i arbetsgruppen som därutöver består av Clas Andersson, Mattias Sandström, Johan Gidlöf samt Christer Lundh alt. Stefan Öberg. Arbetsgruppens uppdrag sträcker sig över perioden 2018-2019.

#### *Aktiviteter under 2018*

1. Varje databas tar fram en lista över samtliga ingående församlingar och hur de är kodade. Underlaget skickas till Mats Berggren senast den 1/3.
2. Probleminventering och analys med utgångspunkt i de underlag som har tagits fram.
3. Arbetet påbörjas med att utarbeta:
  - Gemensamma kodscheman för olika tidsperioder (steg 1)
  - En översättningstabell för geografisk kodning över tid (steg 2)

#### *Tidsplan*

- Probleminventering och analys ska vara klar under 2018.
- Det fortsatta 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 inletts under samordningsprojektet 2016-2017 med databaserna i Lund och Umeå. En ny projektgrupp tillsätts för att leda det fortsatta arbetet. Projektgruppen består av Annika Westberg, Johan Gidlöf, Mats Berggren, Clas Andersson och Luciana Quaranta.

### *Aktiviteter under 2018*

1. En probleminventering och en analys av hur det fortsatta arbetet med IDS-konvertering ska läggas upp utförs under ledning av projektgruppen.
2. En systemdokumentation för omkodningen till IDS tas fram enligt den struktur som finns tillgänglig via EHPS-net.
3. En gemensam workshop anordnas med IDS-experterna George Alter (USA) och Kees Mandemakers (Nederländerna).
4. Riksarkivet, Stockholms stadsarkiv och Umeå besöker Lund för diskussion och auskultation.
5. Databaserna i Umeå och Lund konverteras till IDS v.4.
6. Utveckling av verktyg för konvertering av GOPP- SweCens- och Rotemansdatabaserna.

### *Tidsplan*

- IDS-versioner av databaserna i Lund och Umeå ska finnas tillgängliga vid slutet av 2018.
- Ett verktyg för konvertering av GOPP- SweCens- och Rotemansdatabaserna ska finnas tillgängligt vid slutet av 2018 (Delresultat 2018:a)

## Modul 3

---

### A. Utveckling av ett enkelt uttagsverktyg på webportalen

#### *Nuläge*

Arbetet har ännu inte startat. En arbetsgrupp tillsätts med uppdrag att ta fram riktlinjer för vilken funktionalitet uttagsverktyget ska ha i en första version och för vad som bör ses som ett slutmål. Avsikten är att uttagsverktyget även ska kunna fungera för data som förenade med restriktioner. Maria Larsson är sammankallande i arbetsgruppen som därutöver består av Clas Andersson och Mats Berggren med Johan Gidlöf som bisittare.



*Aktiviteter under 2018*

1. En workshop genomförs under hösten 2018 med syfte att ta fram en kravspecifikation på hur uttagsverktyget bör se ut, i en första version och fullt utvecklat.
2. Utveckling av en första version av uttagsverktyg, i samarbete mellan DDB och Riksarkivet.

*Tidsplan:*

- Ett första uttagsverktyg ska finnas tillgängligt på webportalen vid slutet av 2018 (Delresultat 2018:b)
- En kravspecifikation ska vara upprättad för den funktionalitet uttagsverktyget ska ha i sin slutversion och en plan ska finnas för utvecklingen.

**B. Utveckling av webbportal***Nuläge*

Under samordningsprojektet har en enkel webbsida med information om projektet och de ingående databaserna tagits fram. Denna behöver utvecklas både avseende struktur och innehåll.

*Aktiviteter under 2018*

1. Varje databas ser över den nuvarande informationen och skickar ev. uppdateringar till Maria Larsson senast den 1 mars.
2. Ledningsgruppen arbetar vidare under året med frågan om struktur och innehåll.

*Tidsplan:*

- Nuvarande webbsida ska vara uppdaterad med aktuell information senast under hösten 2018
- Ett plan för vidareutveckling av webbsida/portal ska finnas i slutet av 2018.

## Modul 4

---

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

Arbetet har ännu inte startat. En arbetsgrupp tillsätts med uppdrag att ta fram en kravspecifikation för det länkingsverktyg som ska utvecklas och en plan för arbetet. Maria Larsson är sammankallande i arbetsgruppen som därutöver innefattar Johan Gidlöf, Mats Berggren och Björn Eriksson.

*Aktiviteter under 2018*

1. En workshop genomförs under hösten 2018 med syfte att ta fram en kravspecifikation för det länkingsverktyg som ska utvecklas och en plan för arbetet.

*Tidsplan:*

- En kravspecifikation ska var utarbetad under 2018.
- Utvecklingen av verktyget kommer att ske 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. Arbetet har ännu inte startat.

*Aktiviteter under 2018*

1. Inregistrering av nya data ur församlingsböcker påbörjas.

*Tidsplan*

Datainsamlingen kommer att pågå hela bidragsperioden 2018-2022. Det kommer inte att vara helt avslutat 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. Arbetet har ännu inte startat.

*Aktiviteter under 2018*

1. Förstudie, analys och dokumentation genomförs inför den komplettering som behöver göras

*Tidsplan*

Förstudie, analys och dokumentation är färdigställd under 2018. Kompletteringen av data och länkning genomförs i huvudsak 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.

#### *Aktiviteter under 2018*

1. Inskanning av återstående dödsorsaksintyg i Stadsläkarens arkiv.
2. Upparbetning av tidigare inregistrerade data.
3. Kompletterande registrering av återstående dödsorsaker påbörjas.

#### *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*

Arbetet har inte startat

#### *Aktiviteter under 2018*

1. Inga aktiviteter är planerade under 2018

#### *Tidsplan*

Arbetet kommer att starta först 2019

## 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 2018*

1. Inkomst- och taxeringsuppgifter för 1940, 1945 och 1947 är inregistrerade, länkade och inkluderade i SEDD (Delmål 2018:3)

#### *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 2018*

1. Inga aktiviteter är planerade under 2018.

#### *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 2018*

1. Inskanning av källor från Nysätra församling
2. Förstudie inför kompletteringsarbetet

#### *Tidplan*

Skanning och förstudie avslutas under 2018. 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.

#### *Aktiviteter under 2018*

1. En förstudie genomförs inför utvecklingen av ett nytt inregistreringsformulär till mjukvaran PERSONA, avsett för insamling av ekonomiska data i inkomst- och taxeringslängder.

#### *Tidplan*

Förstudien avslutas under 2018.

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.

#### *Aktiviteter under 2018*

1. Utveckling av ett inregistreringsformulär till mjukvaran PERSONA avsett för insamling av folkräkningsdata.

#### *Tidplan*

Ett nytt formulär ska finnas tillgängligt under 2018.

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. Riksarkivet har tecknat ett samarbetsavtal med Arkiv Digital som gör det möjligt att få tillgång till en kopia för forskningsändamål, som i upparbetad form kommer att inkluderas i SweCens-databasen och länkas till övriga folkräkningar.

#### *Aktiviteter under 2018*

Kodning av yrken genomförs.

#### *Tidplan*

Upparbetningen ska var genomförd under 2019.

# 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 2018*

Inga aktiviteter är planerade under 2018

### *Tidplan*

Arbete inom denna modul kommer att starta först 2020.