

SwedPop – Svenska befolkningsdatabaser för forskning

Verksamhetsberättelse 2024

Sammanfattning

Under 2024, det andra året inom innevarande bidragsperiod, har arbetet inom modulerna i huvudsak fortsatt följt upprättad tidsplan. Några smärre förseningar föreligger, vilka inte bedöms påverka den övergripande planeringen. Nya datamängder tillgängliggjordes för forskning när en ny datarelease publicerades under januari på [swedpop.se](https://doi.org/10.48524/SWEDPOP-V3-0) <https://doi.org/10.48524/SWEDPOP-V3-0>. Arbetet inom de gemensamma modulerna **M1-M4** fokuserade på slutdokumentation av harmoniseringsarbetet, utveckling av ny mjukvara samt länkning av nya och tillkommande data. Inom de övriga modulerna pågick arbete med strategiska kompletteringar av de ingående databaserna med syfte att öka infrastrukturens vetenskapliga värde, samt med att tillföra nya datamängder till SwedPop. Avvikelser i förhållande till upprättad budget förelåg i några moduler, främst föranledda av att kostnader har förskjutits i förhållande till budgetår. Med enstaka undantag var avvikelserna av marginell karaktär. Nyckeltalen pekar på en ökad användning av infrastrukturen under 2024 och på en förbättrad könsbalans mellan manliga och kvinnliga användare. En ökad satsning på bred marknadsföring till akademiska användare kommer att göras när det nya outputformatet finns på plats under 2025. Under året har ett flertal presentationer av infrastrukturen gjorts, både nationellt och internationellt. Verksamhetsberättelsen har tagits fram av infrastrukturens föreståndare.

Populärvetenskaplig beskrivning

Den nationella forskningsinfrastrukturen SwedPop tillgängliggör harmoniserade och länkade befolkningsdata på individnivå för forskning. Kärnan utgörs av sex stora forskningsdatabaser: DDB vid Umeå universitet, SEDD vid Lunds universitet, GOPP vid Göteborgs universitet, SweCens vid Riksarkivet och Rotemansdatabasen vid Stockholms stadsarkiv. Sveriges Dödbok tillgängliggörs i samarbete med Sveriges Släktforskarförbund. Under innevarande bidragsperiod kommer även migrantdatabasen EMIWEB att tillgängliggöras för forskning. Tillsammans erbjuder databaserna sammanhängande information om den svenska befolkningen c. 1700–1968, med regionala nedslag, däribland städerna Stockholm och Göteborg. Data som inte innehåller personuppgifter kan fritt laddas ned från <https://swedpop.se/>, medan data som omfattas av GDPR tillgängliggörs efter särskild prövning. Utifrån källornas noggranna noteringar om födelse, död, bostadsort, yrke, familjebildning, flyttningar, familjeförhållanden, och ekonomisk information kan forskare studera individer, familjer och större befolkningar i sin sociala och ekonomiska kontext över tid, rum och generationer. Möjligheten att länka data till moderna befolkningsregister hos SCB ger förutsättningar för flergenerationsstudier som inkluderar nuvarande befolkning.

Beskrivning av verksamheten

Tidsplan

Sammantaget kan det konstateras att arbetet inom infrastrukturens moduler, med några få undantag, har följt upprättad tidsplan och budget och att de smärre förseningar som förelåg vid slutet av 2024 sannolikt inte kommer att ha någon större påverkan på de långsiktiga målsättningarna. Då förändringarna har varit marginella har GANTT-schemat inte behövt uppdaterats.

- Releasen av det nya outputformatet i uttagsverktyget har försenats med 1–2 månader i förhållande till tidsplanen och planeras nu äga rum under första kvartalet 2025.
- Användarworkshopen med fokus på datauttag, statistiska metoder och programvara har medvetet skjutits upp till dess det nya outputformatet är på plats och kommer att genomföras under senare delen av 2025.
- Standardiseringen av utökade geografiska variabler har försenats, då arbetet visade sig vara mer komplext än förväntat. Det kommer att slutföras senast 2026.

En viss oro har förelegat avseende datainsamlingen i M13 *Inkludering av Folkräkningen 1940 i SwedPop*, men utifrån vidtagna åtgärder är bedömningen att dataleverans kommer att kunna ske inom ramen för upprättad tidplan (se vidare under rubriken Riskanalys).

Konstruktion, utveckling och drift av infrastrukturen

Arbetet i SwedPop bedrivs inom 11 moduler som tillsammans bidrar till att uppnå de vetenskapliga målen för infrastrukturen. De gemensamma modulerna 1–4 omfattar projektledning/organisation och administration (M1), standardisering, kodning och datastruktur (M2), utveckling och underhåll av uttagsverktyg samt användarstöd (M3) och länkning, utvärdering och dokumentation (M4). M1-4 är starkt beroende av varandra och arbetet inom dem är av avgörande betydelse för att SwedPop ska bli en välfungerande resurs för svensk forskning. Återstående sju moduler är partsspecifika. Arbetet inom dessa moduler består huvudsakligen av strategiska kompletteringar av befintliga databaser med syfte att öka infrastrukturens vetenskapliga värde, samt tillförandet av nya datamängder till SwedPop.

M1 har en stödjande roll i projektet. Här finns budget för projektledning, administration, resor och projektmöten. Inga specifika verksamhetsmål har utarbetats för 2024.

Inom **M2** fortsatte arbetet med slutdokumentation av den omfattande harmonisering och standardisering av kodning och datastruktur till internationell kodstandard, som genomförts och implementerats inom SwedPop. Målsättningen är att all dokumentation ska finnas öppet tillgänglig på swedpop.se under 2025. Öppen tillgång till funktionella svenska versioner av internationell kodstandard för centrala variabler i befolkningsforskning har ett stort vetenskapligt värde i sig, och bidrar bland annat till att underlätta implementeringen av FAIR-principerna inom forskningsområdet.

Aktiviteterna inom **M3** har under året rymt kontinuerliga uppdateringar av swedpop.se och uttagsverktyget. Inom denna modul har även det nya outputformatet för data utvecklats, som lanseras under våren 2025 och kommer att göra det smidigare för forskare att ladda ned och analysera SwedPop-data. Formatet är anpassat till de vanligaste analysprogrammen: Stata, SAS, SPSS och R. Ett formulär för beställning av skyddade data har också utvecklats som kommer att implementeras i uttagsverktyget under 2025–26. Formuläret möjliggör en säker och effektiv bearbetning av beställningar som kräver utlämningsprövning. I januari 2024 publicerade en ny datarelease, <https://doi.org/10.48524/SWEDPOP-V3-0>.

Inom **M4** genomfördes löpande länkning mellan de olika databaserna efter gemensamma principer inklusive dokumentation. Under 2024 har en fullständig länkning gjorts mellan alla tillgängliga folkräkningar 1880–1910 och Sveriges Dödbok. Länkarna kommer att publiceras på swedpop.se med open access.

Aktiviteterna inom **M5, M8-M12, M15** har avsett datainsamling samt kompletteringar av databaserna GOPP, Rotemannen, SEDD, DDB och SweCens. Arbetet har i huvudsak följt fastställd budget och verksamhetsplan. I några fall har det av praktiska skäl gjorts en omprioritering av den ordning och omfattning i vilken informationen i källorna har inregistrerats i relation till den plan som beskrevs i ansökan. Inom **M13** har alternativa vägar sökts för att säkerställa datainsamling inom ramen för gällande budget.

Samverkan med andra infrastrukturer

Under 2024 har samarbete och erfarenhetsutbyte mellan nationella forskningsinfrastrukturer fortsatt och utvecklats inom ramen för Infraforum. Ett antal fysiska och digitala möten har genomförts för att diskutera gemensamma angelägenheter. Till höstens möte hade bland annat Maria Wästfelt från Regeringskansliet bjudits in för att informera om den nya regleringen för forskningsdatabaser. Samarbetet med närliggande infrastrukturer i de

nordiska länderna har fortsatt. Under hösten 2024 besökte delar av ledningsgruppen projektgrupperna i Oslo och Tromsö som arbetar med att ta fram Historisk befolkningsregister 1800–1964 (HPR), för erfarenhetsutbyte rörande metoder för ATR-digitalisering. Diskussioner har även förts med projektet LinkLives och danska Rigsarkivet som planerar ett liknande projekt i Danmark. SwedPop har fortsatt varit en aktiv part i arbetet med att implementera en ny internationell kodstandard för dödsorsaker i historiska data, som leds av Radboud University i Nederländerna. Samarbete har fortsatt funnits med amerikanska IPUMS vid University of Minnesota, rörande tillgängliggörande av svenska folkräkningar.

Datahantering och behov av stödjande e-infrastruktur

Under året har frågor kring rättslig grund för hantering av personuppgifter i databaserna i SwedPop utretts och klargjorts. Tillsammans med övriga parter i Infraforum har frågan om behovet av långsiktig rättslig reglering av forskningsdatabaser fortsatt aktualiserats, föranlett av regeringens förslag och senare beslut om långsiktig reglering av vissa forskningsdatabaser.

Förändringar av organisation

Inga organisatoriska förändringar har skett under året.

Styrgruppens arbete

Under 2024 sammanträdde styrgruppen vid två tillfällen: 2024-03-26 och 2024-10-17. Vid det första mötet fastställdes årsredovisningen till VR och juridiska aspekter på tillgängliggörande av data och strategier för breddning av SwedPop:s användarbas diskuterades. Årets andra möte ägnades i huvudsak åt strategiska framtidsytande frågor i anslutning till en kommande ansökan, samt diskussion om VR:s samordningsuppdrag som ska redovisas 2025. Under året har styrgruppen därtill fått information om pågående arbete, metodutveckling och användning av infrastrukturen samt om aktuell forskning möjliggjord av SwedPop-data.

Ekonomiskt utfall

Under 2024 har arbetet inom de olika modulerna finansierats av VR-bidrag och samfinansiering. Data tillgängliggörs fritt via webbportalen swedpop.se varför inga användaravgifter tillämpas. Därav har samfinansieringens omfattning inte påverkats i relation till fastställd budget.

Under 2024 förelåg några avvikelser från budget, vilka framför allt är kan hänföras till att kostnader har blivit förskjutna i förhållande till budgetår. Anledningarna till detta kan i huvudsak sökas i logistiskt relaterade omprioriteringar vid infrastrukturens noder, ofta i relation till tillgänglig kompetens och personalresurser. Inom **M1** förelåg ökade kostnader för ledning. Ett ökat behov av samordning har funnits under året, framför allt relaterat till inkluderingen av en tillkommande nod i konsortiet. Ej budgeterade kostnader för nätverksmöten och erfarenhetsutbyte med andra infrastrukturer inom samma fält har också tillkommit.

Den största differensen gentemot budget föreligger inom **M2** och **M3** där medelförbrukningen har varit högre än budgeterat. Detta beror dels på att aktiviteter som under 2023 försenats till följd av brist på systemutvecklarresurser har genomförts under året, dels på att systemutvecklingskostnader budgeterade till 2025 har tidigare lagts. Differenser ses även i modulerna **M2**, **M8**, **M9** och **M10** beroende på omfördelning av resurser mellan budgetår. Då det under en tid funnits en oro över en allt för låg produktionstakt inom **M13**, sattes ökade resurser in i början av 2024 för att om möjligt påskynda digitaliseringen. Senare under året kunde tillgången till data säkras via en alternativ lösning (se avsnittet Riskanalys).

Inom **M2**, **M5**, **M12** och **M15** var medelsförbrukningen lägre än budgeterat, då vissa aktiviteter av logistiska skäl har fått skjutas fram till 2025. Detta har fått till följd att samfinansieringen under året har understigit 50% i tre av noderna, UMU, LU och RA, vilket kommer att justeras under 2025 och 2026.

Kommentar på nyckeltal

Under 2024 uppmärksammades att det under ett par år beklagligtvis har inrapporterats felaktiga och för höga siffror avseende den internationella användningen av SweCens-data via IPUMS. Detta beror uteslutande på ett missförstånd angående form och rutiner för rapportering av statistik. Detta har nu klarats ut och 2023 års nyckeltal har justerats till korrekta nivåer. Trots detta kan vi konstatera att användningen av SwedPop-data har ökat stadigt för varje år, både sett till antalet användare och antalet projekt. Mellan 2023 och 2024 ökade antalet projekt med +30, antalet användare med +35 och antalet levererade datauttag med +130. Tidigare erfarenhet visar att ökningen av användarbasen har varit särskilt stor i samband med publicering av nya data, varför vi kan anta att tillgängliggörandet av Rotemansdatabasen och Folkräkningen 1950 särskilt har bidragit till ökningen under 2024. Fördelningen av användare mellan de olika användarkategorierna var relativt oförändrad i jämförelse med 2023, med en fortsatt hög andel användare vid andra svenska lärosäten, samt internationella användare. Andelen övriga användare var fortsatt marginell. Antalet anställda inom infrastrukturen har ökat något, både sett till antalet individer och i FTE.

Kommentar på publikationslista

Rapporteringen av publikationer från öppet tillgängliga forskningsinfrastrukturer bygger på att forskare åiterrapporterar eller taggar sina publikationer på ett sätt så att de kan identifieras och inkluderas i redovisningen. Trots att vi begär att forskare som använder SwedPop ska citera data på det sätt som anges i *Terms and Conditions*, med referens till doi så görs det sällan. Detta innebär att den publikationslista som redovisas fortsatt framför allt baseras på de större dataleveranser som sker utanför swedpop.se och som vi därmed har bättre kontroll på. Av samma skäl finns en överrepresentation av publikationer från parterna i konsortiet. Inom SwedPop arbetar vi hela tiden med att hitta vägar till att stimulera användare att göra sina publikationer sökbara, så att det ska bli möjligt att på ett mer tillförlitligt sätt mäta infrastrukturens impact i form av vetenskapliga publikationer. Inför redovisningen av år 2024 har 33 publikationer baserade på SwedPop-data identifierats. Därutöver utgör SwedPop-data en central del i ERC-projektet *Relative Health: Long-Run Inequalities in Health and Survival Between Families and Across Generations* (PI: Ingrid Van Dijk), vilket ännu inte har generat några publikationer. Antalet publikationer minskade något jämfört med 2023 (39) men med tanke på svårigheterna att få fram en tillförlitlig statistik går det inte att dra några långtgående slutsatser utifrån detta.

Jämställdhet

Jämställdhetsarbetet inom SwedPop kännetecknas av en långsiktig strävan att uppnå de mål som anges i upprättad jämställdhetsplan. Könslans har i möjligaste mån beaktats när arbetsgrupper har utsetts och det praktiska arbetet vid de olika noderna är fortsatt förhållandevis jämställt. Under 2024 utjämnades könslansen i ledningsgruppen ytterligare, när en ny ledamot (kvinna) utsågs. Både kvinnor och män har medverkat vid presentationer av infrastrukturen. Vi kan notera att könslansen bland infrastrukturens användare har fortsatt att utjämnas, även om det fortsatt finns utrymme för förbättring. Under 2024 var 35% av användarna kvinnor, vilket kan jämföras med 30% 2023 (OBS baserad på reviderad statistik från 2023). Könslansen bland de anställda vid infrastrukturen var fortsatt god. Under 2024 var 23 män och 19 kvinnor anställda, vilket motsvarar 7,15 respektive 8,36 FTE. Infrastrukturens föreståndare är kvinna och bland teknisk-administrativ personal anställd vid infrastrukturen fanns en stor kvinnlig övervikt, framför allt sett till FTE. Det praktiska arbetet

inom olika arbetsgrupper var fortsatt förhållandevis jämförbart, även om könsbalansen minskade något under året till följd av personalförändringar. Styrgruppen hade en jämn könsfördelning.

Risikanalyt

Ledningsgruppen följer löpande upp arbetets fortskridande i de olika modulerna gentemot verksamhet och budget. Inga nya risker har identifierats under året. Däremot har fortsatt uppmärksamhet ägnats åt arbetet inom modul 13, med syfte att säkerställa att folkräkningen 1940 kan tillgängliggöras innan bidragsperiodens slut. Under 2023 gjordes en överenskommelse med amerikanska släktforskningsföretaget My Heritage avseende leverans av data till SwedPop, men det formella avtalsskrivandet kom att dra ut på tiden till följd av juridiska komplikationer och under stora delar av 2024 var det oklart ifall något avtal skulle kunna komma till stånd. Under senhösten 2024 nåddes dock en juridisk godtagbar lösning, som innebär att data kommer att tillställas SwedPop inom ramen för nuvarande bidragsperiod, enligt tidigare överenskommelse

Utbildningsinsatser, outreach

Under 2024 har presentationer av SwedPop gjorts vid följande tillfällen:

- 2024-02-29 Seminarium University of Iceland, Reykjavík, presentation med titel *Sharing data with the public: experiences from Sweden*.
- 2024-03-18 Möte med den nationella infrastrukturen InfraVis.

Metodinriktade presentationer som rör infrastrukturens arbete har även gjorts vid konferenser och workshops nationellt och internationellt. Allt fler resurser för användarstöd (user guide, dokumentation och FAQ) har publicerats på swedpop.se under året. En viss ökning av efterfrågan på användarstöd har noterats i samband med publicering av folkräkningen 1950 för forskning.

Övrigt

I juni 2024 publicerades resultatet från behovsinventeringen av nationell forskningsinfrastruktur 2023–2024 där SwedPop bedömdes fortsatt ha hög vetenskaplig relevans och vara av nationellt intresse. Inför den kommande utlysningen föreslog RFI att en samordning skulle ske med ATLASS, en ny resurs med skoldata och demografisk information på individnivå. Stora delar av hösten 2024 ägnades åt överläggningar med ATLASS för att tydliggöra förutsättningarna för samverkan, där ATLASS utgör en ny nod i SwedPop, samt för att förbereda en ansökan om bidrag till infrastruktur av nationellt intresse i den utlysning som öppnade i december 2024.