



Riksskivets

# SwedPop

**Svenska befolkningsdatabaser för forskning**

NORDIG 2017, Mats Berggren, Riksskivets, Sverige

# SwedPop

- SwedPop: En forskningsinfrastruktur
- SweCens: Folkräkningar för forskning
- SwedPop: Organisation och arkitektur

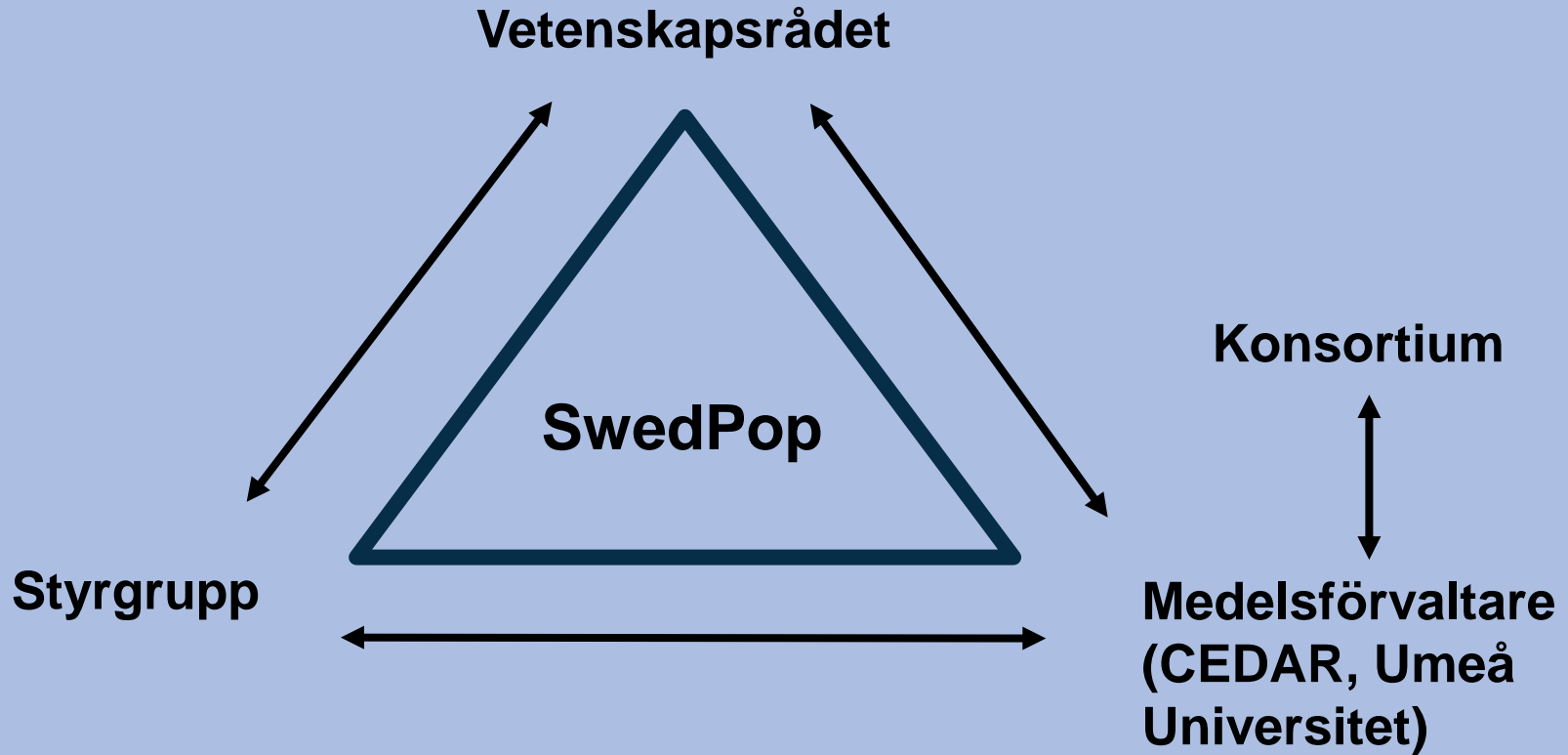
# SwedPop

- SwedPop: En forskningsinfrastruktur
- SweCens: Folkräkningar för forskning
- SwedPop: Organisation och arkitektur

# SwedPop: en forskningsinfrastruktur

- **Forskningsinfrastruktur**
  - Vetenskapsrådet (Swedish Research Council) finansierar
  - En infrastruktur måste utgöras av ett konsortium bestående av minst tre lärosäten
  - Finansiering ges för upp till åtta år
  - Utlysning vartannat år
  
- **SwedPop**
  - Besked i september 2017, finansiering i minst fem år fr.o.m. 2018
  - Bidrag: 6-8 miljoner per år för perioden 2018-2022, medfinansiering 50%
  - Riksarkivets del: 1.7-2.5 miljoner per år 2018-2022, medfinansiering 50%
  - En ny tjänst fr.o.m. 2018, flera på sikt
  - Ny roll för Riksarkivet

# Rollfördelning – ansvar för nationell infrastruktur



# Medlemmar i SwedPop-konsortiet

- **Umeå universitet, Centrum för demografi och åldrandeforskning (CEDAR)**
  - Databaserna POPUM, POPLINK och TABVERK
  - POPUM och POPLINK omfattar tillsammans cirka en miljon personposter för åren 1680-1953
- **Lunds universitet, centrum för ekonomisk demografi**
  - Scania Economic-Demographic Database (SEDD)
  - Omfattar cirka 100000 personposter för åren 1646-1968
- **Göteborgs universitet, institutionen för ekonomi och samhälle**
  - Gothenburg Population Panel (GOPP)
  - Omfattar 12500 observationer av 3767 personer för åren 1915-1943
- **Riksarkivet**
  - SweCens-databasen innehållande folkräkningarna 1880, 1890, 1900, 1910
  - Omfattar 20.2 miljoner personposter och 5.5 miljoner hushållsposter
  - Folkräkningarna 1930, 1940 och 1950 kommer att tillföra 6.1, 6.5 och 7 miljoner personposter
- **Stockholms stadsarkiv**
  - Rotemansdatabasen
  - Omfattar 6.5 miljoner observationer av 2 miljoner personer för åren 1878-1926

# SwedPop: mål

- **Dataregistrering**
  - Fortsatt registrering av individdata. Registrering av ekonomiska förhållanden samt dödsfall och dödsorsaker
- **Skapa gemensam datastruktur**
  - Överföra informationen i samtliga ingående databaser till den nya standarden "Intermediate Data Structure" (IDS) för att möjliggöra uttag av jämförbara dataset från samtliga databaser
- **Dokumentation**
  - Samtliga databaser är beskrivna och dokumenterade, men det behövs ett arbete för att koordinera dokumentationen och få den enhetligare för de ingående databaserna
- **Harmonisering av kodning**
  - Det krävs en harmonisering av kodningsmetoder och kodlistor, exempelvis för geografisk kodning och yrkeskodning, för att göra de ingående databaserna mer enhetliga och mer jämförbara för forskare
- **Gemensam e-infrastruktur**
  - Det ska skapas en gemensam webbportal för SwedPop så att information om samtliga databaser kan fås från ett ställe. Via webbportalen ska det vara möjligt att välja variabler och beställa datauttag från samtliga databaser. Det ska också vara möjligt att från webbportalen direkt ladda ner data som inte omfattas av sekretess
- **Gemensam utveckling**
  - I fortsättningen ska nya registreringsinsatser och andra initiativ koordineras inom konsortiet



# SwedPop

- SwedPop: En forskningsinfrastruktur
- SweCens: Folkräkningar för forskning
- SwedPop: Organisation och arkitektur



# Svensk folkbokföring genom tiderna

- **Kyrkolagen 1686**
  - Födelse-, dop- och vigsellängder
  - Husförhörslängder
- **Tabellverket 1749**
  - Aggregerad data
  - Sockentabeller i kyrkoarkiven
  - Sammandrag och statistik i Tabellkommissionens arkiv
  - Registrerat och publicerat av CEDAR vid Umeå universitet:  
<http://www.cedar.umu.se/ddb/ships>
- **Statistiska centralbyrån (SCB) 1860**
  - Folkräkningar 1860 – 1945 baserade på husförhörslängder och församlingsböcker (1895-1945)
  - Folkräkning baserad på mantalslängden 1950
  - Folk- och bostadsräkningar 1960 – 1990
  - Summariska folkmängdsredogörelser 1860 – 1972
  - Hushålls- och bostadsräkningen 2011 (Census 2011)

# Riksarkivets registrering av folkräkningar

- **Registreringen administreras av SVAR**
  - Registreringen startade 1994 med folkräkningen 1890, projektet ALU 1000
  - För närvarande 10 arbetsplatser
  - Total arbetsstyrka cirka 70 personer
  - Registrerade kompletta folkräkningar: 1880, 1890, 1900 och 1910
  - För närvarande registreras folkräkningen 1930
  - Registrering sker lokalt. Datafiler samlas in av SVAR och importeras till en SQL-databas
  - Folkräkningarna sökbara via Riksarkivets söktjänst (gratis från 2018):  
<http://sok.riksarkivet.se/folkrakningar>
- **Organisation och finansiering**
  - Löne-medel från kommunerna och från arbetsförmedlingen
  - Arbetslösa och arbetshandikappade
  - Lokaler och utrustning från kommunerna
  - Källmaterial (skannade bilder), programvara, utbildning, arbetsledning och support från SVAR i Ramsele

# SweCens: Folkräkningar för forskning

- **SweCens I (2011-2012)**
  - Finansierat av ett planeringsbidrag från Vetenskapsrådet
  - Finansiering även från NAPP-projektet i USA
  - Kodning av folkräkningarna 1880, 1890 och 1900 för leverans till NAPP-projektet vid Minnesota Population Center (MPC), University of Minnesota, USA
  - Utvärdering av metoder för länkning av de svenska folkräkningarna
- **SweCens II (2013-2015)**
  - Finansiering av ett bidrag för ”stora databaser” från Vetenskapsrådet
  - Länkning av folkräkningarna 1880, 1890 och 1900
  - Leverans av sex länkade samples till MPC (Leverans 2014)
  - Kodning av folkräkningen 1910 för leverans till MPC (Leverans 2016)
  - Länkning av folkräkningen 1910 till folkräkningarna 1880-1900 (Under 2017)
  - Digitalisering, registrering, kodning och länkning av folkräkningen 1930
  - Folkräkningarna 1880, 1890, 1900 och 1910 är tillgängliga via MPC:s webbplats:  
<http://www.nappdata.org>

# Svenska historiska folkräkningar

Folkräkning:	Skannad:	Registrerad	Ingår i NAPP
År 1860	Ja	Jämtland (63614 personposter)	Nej
År 1868 Stm, Gbg	Nej	Nej	Nej
År 1870	Ja	Norrbottn och Västerbotten (167948)	Nej
År 1880	Ja	Hela Sverige komplett (4624825)	Ja, sedan 2015
År 1890	Ja	Hela Sverige komplett (4846123)	Ja, sedan 2012
År 1900	Ja	Hela Sverige komplett (5200111)	Ja, sedan 2008
År 1910	Ja	Hela Sverige komplett (5586353)	Ja, sedan 2017
År 1920	Ja	Nej	Nej
År 1930	Ja	Registrering pågår inom SweCens	Nej
År 1935-1936	Nej	Nej	Nej
År 1940	Nej	Nej	Nej
År 1945	Nej	Nej	Nej
År 1950	(Ja)	(Registrerad 2015 av Arkiv Digital)	Nej

# Svenska folk- och bostadsräkningar

## Folk- och bostadsräkningar (born digital):

Folk- och bostadsräkningen 1960

Folk- och bostadsräkningen 1965

Folk- och bostadsräkningen 1970

Folk- och bostadsräkningen 1975

Folk- och bostadsräkningen 1980

Folk- och bostadsräkningen 1985

Folk- och bostadsräkningen 1990

## Census 2011 (Hushålls- och bostadsräkningen 2011)

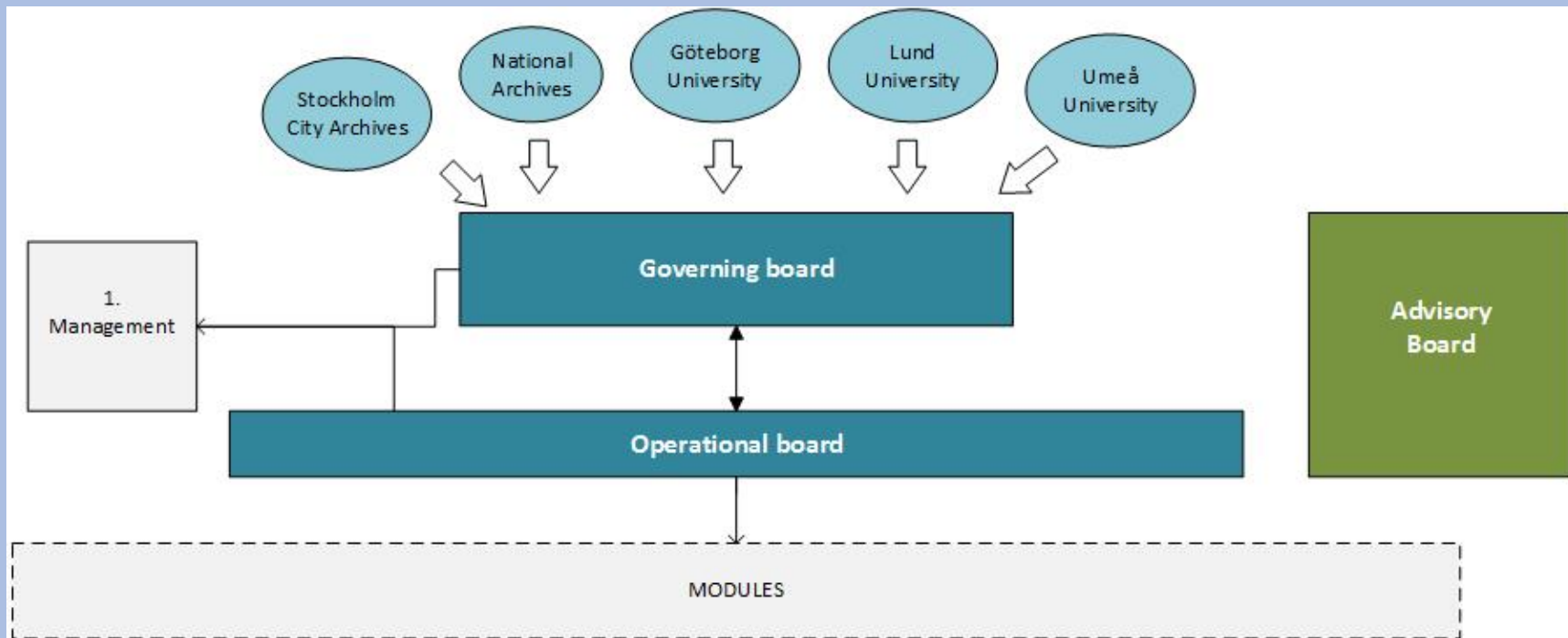
### Information på SCB:s webbplats:

<http://www.scb.se/hitta-statistik/statistik-efter-amne/befolkning/befolkningens-storlek-och-forandringar/hushalls-och-bostadsrakning-census>

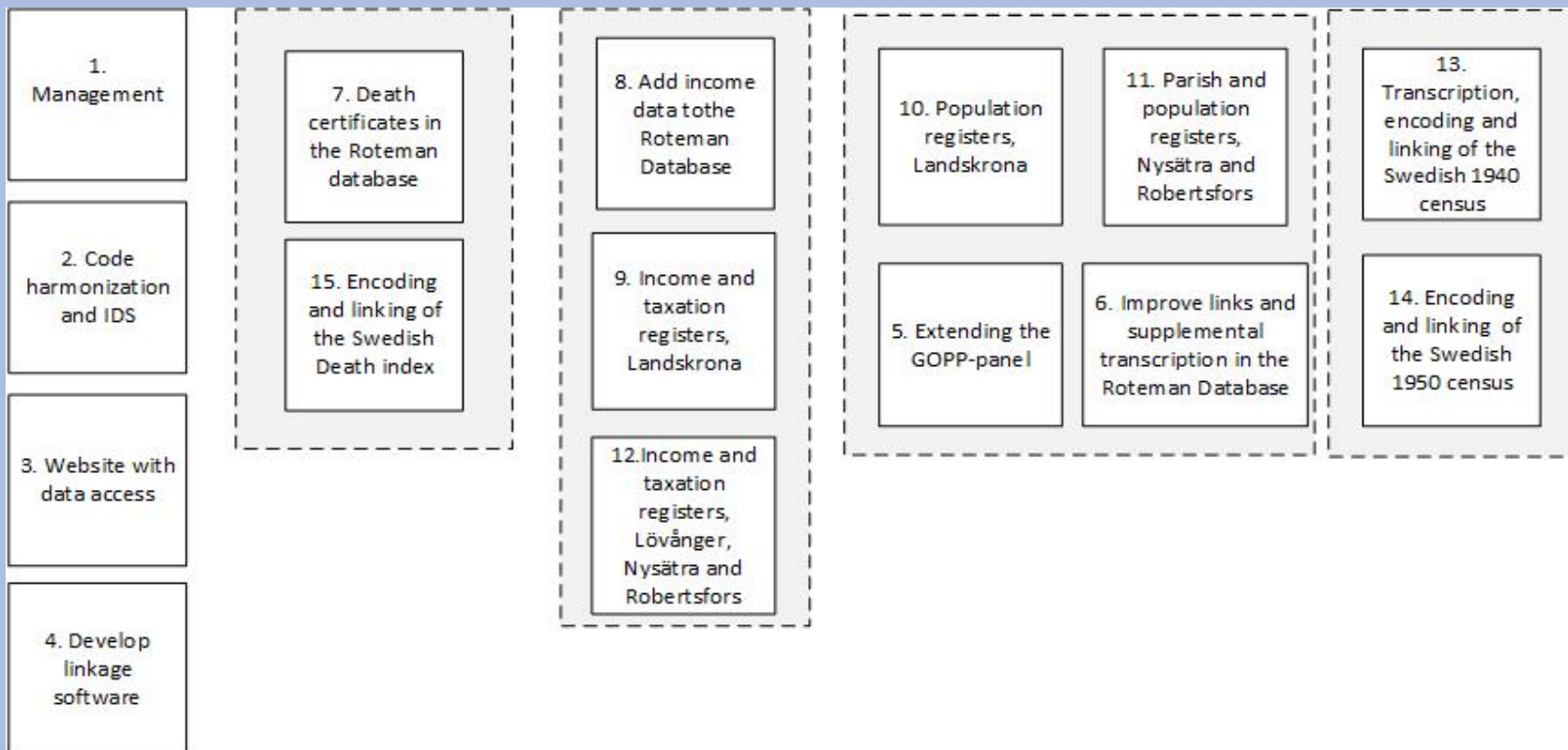
# SwedPop

- SwedPop: En forskningsinfrastruktur
- SweCens: Folkräkningar för forskning
- SwedPop: Organisation och arkitektur

# SwedPop: Organisation



# SwedPop: Moduler





# Gemensamma moduler

- **Module 1: Management**
  - Gemensam administration av forskningsinfrastrukturen
- **Module 2: Code harmonization and IDS**
  - Harmonisering av koder mellan de olika databaserna, geografiska koder, yrkeskoder (HISCO)
  - Överföring av data till formatet "Intermediate data structure" (IDS)
- **Module 3: Expanding the Swedpop webportal with tools for data access**
  - Gemensam webbportal i Umeå
  - Gemensamma beställningsrutiner för data
  - Webbapplikation för att möjliggöra uttag och nedladdning av data online för data som inte omfattas av sekretess
- **Module 4: Developing effective software for a high-quality linkage and linking the core databases together**
  - Länkningsprogramvara ska utvecklas.
  - Länkning av personer mellan samtliga ingående databaser

# Moduler för dödsfall och dödsorsaker

- **Module 7: Transcribe and link death certificates 1878-1926 in the Roteman Database**
  - Registrering av dödsbevis från hälsovårdsnämndens och Stadsläkarens arkiv
  - Genom att knyta dödsbevisen till personerna i databasen Rotemannen ökar användbarheten för medicinsk och samhällsvetenskaplig forskning
- **Module 15: Encoding and linking of the Swedish Death Index**
  - Riksarkivet har träffat en överenskommelse med Sveriges släktforskarförbund om att få tillgång till databasen "Sveriges dödbok" för vetenskapligt bruk
  - Dödboken är en databas som kommer att omfatta samtliga avlidna från 1860 till nutid
  - Genom att länka personerna i dödboken till folkräkningarna och de övriga databaserna inom SwedPop öppnas många nya möjligheter för forskning

# Moduler för ekonomisk information

- **Module 8: Add economic data to the Roteman Database**
  - Rotemansdatabasen kommer att kompletteras med ekonomiska uppgifter från Kronotaxeringslängderna för 1901 och 1910 och med uppgifter om bostäder och hyror för 1901
- **Module 9: Income and taxation registers in Landskrona 1905-1947**
  - Uppgifter om inkomster ur taxeringslängderna kommer att registreras för Landskrona stad för perioden 1905-1947
- **Module 12: Income and taxation registers, Löfvånger, Nysätra and Robertsfors**
  - Uppgifter om inkomster ur taxeringslängderna kommer att registreras för vart fjärde år från 1881 till 1950 för de tre Västerbottensförsamlingarna Löfvånger, Nysätra och Robertsfors
  - Detaljerad individuell ekonomisk information från Skåne, Stockholm och Västerbotten kommer att ge nya möjligheter till forskning inom samhällsvetenskap och ekonomisk historia

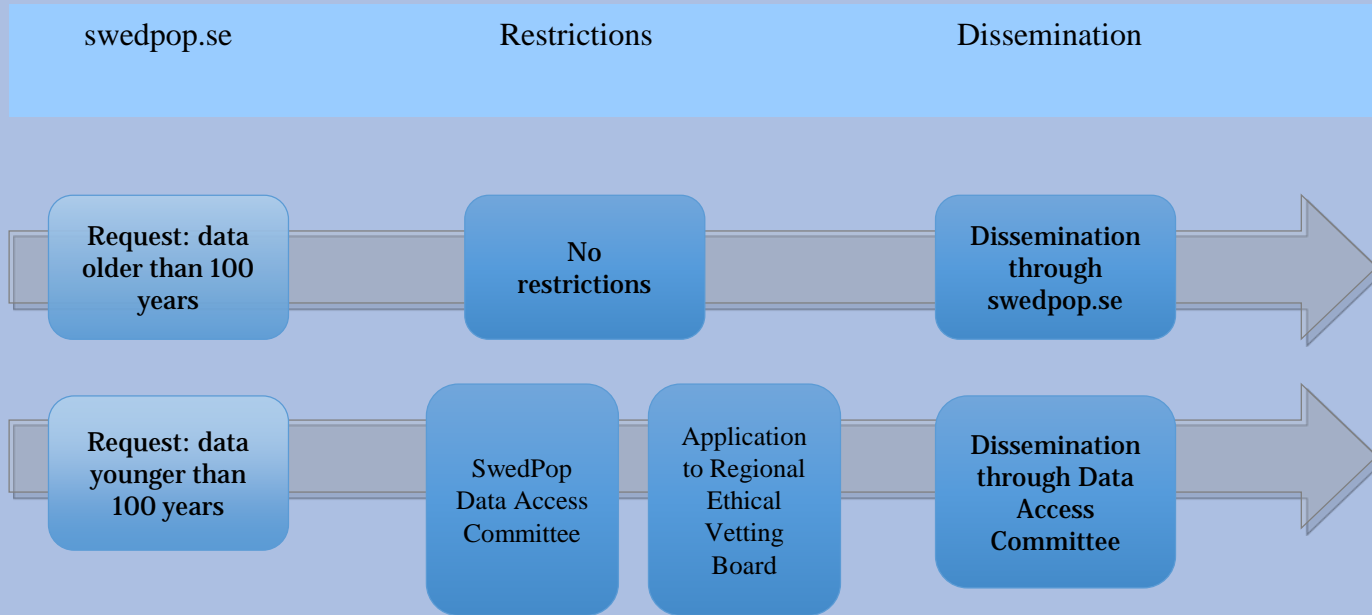
# Moduler för kompletterande individregistrering

- **Module 5: Extending the GOPP panel**
  - GOPP kommer att utökas med data för vart fjärde år från 1947 till 1967
- **Module 6: Improved links and supplementary transcription in the Roteman Database**
  - Databasen kommer att kompletteras med geografisk kodning och yrkeskodning
  - Den interna länkningen som knyter samman observationer för personer kommer att förbättras
- **Module 10: Population registers Landskrona 1882-1910**
  - Databasen kommer att kompletteras med registrering av data ur husförhörslängder, församlingsböcker, födelselängder, vigsellängder och dödböcker för perioden 1882-1910
- **Module 11: Parish and population registers, Nysätra and Robertsfors**
  - Registrering av information ur husförslängder, församlingsböcker, födelseböcker, vigselböcker och dödböcker för perioden 1700-1950
  - Denna registrering kommer geografiskt att knyta samman det registrerade klustret av församlingar kring Skellefteå med området runt Umeå.

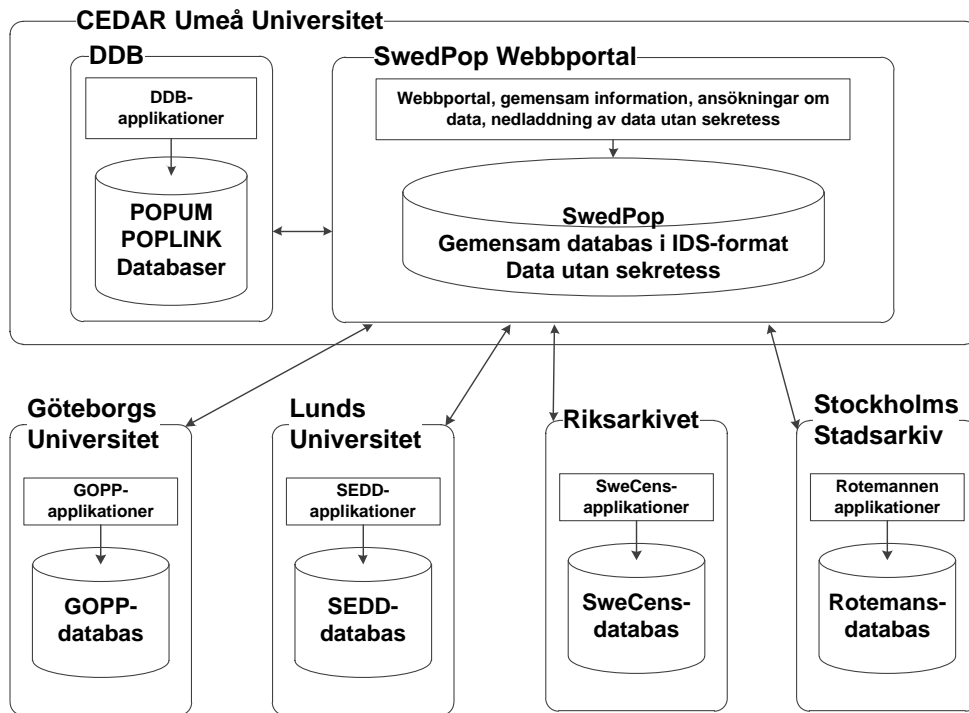
# Moduler för folkräkningar

- **Module 13: Transcription, encoding and linking of the Swedish 1940 census**
  - Så snart folkräkningen 1930 är färdigregistrerad kommer registrering av folkräkningen 1940 att påbörjas
  - I takt med att registreringen genomförs kommer folkräkningen att kodas för vetenskapligt bruk och länkas till övriga folkräkningsår och till de övriga databaserna inom SwedPop
- **Module 14: Encoding and linking of the Swedish 1950 census**
  - Riksarkivet har träffat en överenskommelse med företaget Arkiv Digital om att få tillgång till folkräkningen 1950 för vetenskapligt bruk
  - Folkräkningen 1950 kommer att kodas på samma sätt som övriga folkräkningar och länkas både till tidigare folkräkningar och till övriga databaser inom Swedpop
- **(The Swedish 1860 census)**
  - Folkräkningen 1860 ingår egentligen inte i Swedpop-projektet.
  - Diskussioner förs med Sveriges Släktforskningsförbund om en eventuell registrering av denna folkräkning på frivillig basis. I så fall skulle Riksarkivet kunna få tillgång till databasen för vetenskapligt bruk

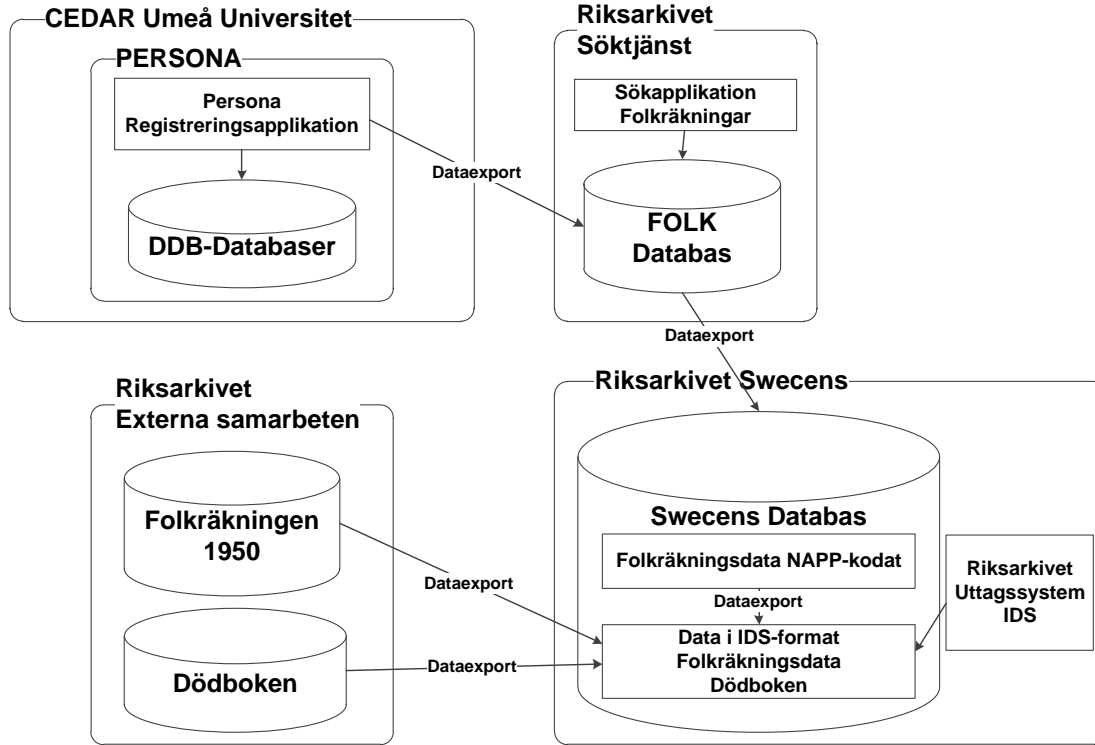
# SwedPop: Dataaccess



# SwedPop: Arkitektur



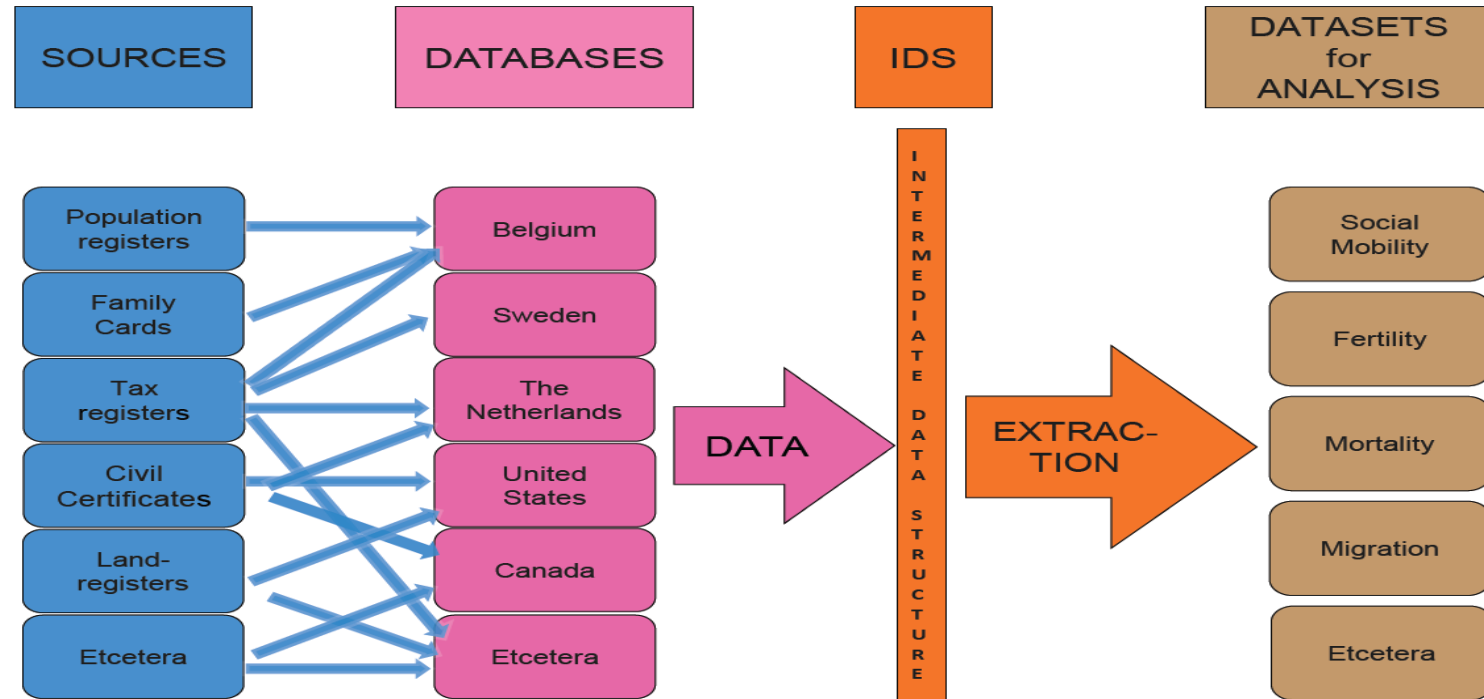
# SwedPop: Arkitektur SweCens





# Intermediate Data Structure (IDS)

Figure 1 *Strategy with intermediate structure collecting data for scientific research from historical longitudinal databases.*



# IDS Datamodell

**Table INDIV**

Id	Id_D	Id_I	Source	Type	Value	Value_Id_C	Timestamp
1	SweCens	109621556	Census	Last_Name	Öhrman		1900-12-31
2	SweCens	109621556	Census	First_Name	Hanna Viktoria		1900-12-31
3	SweCens	109621556	Census	Birth_Date	1874		1900-12-31
4	SweCens	109621556	Census	Birth_Location		248104000	1900-12-31
5	SweCens	109621556	Census	Sex	Female		1900-12-31
6	SweCens	109621556	Census	Civil_Status	Married		1900-12-31

**Table INDIV\_INDIV**

Id	Id_D	Id_I_1	Id_I_2	Source	Relation
1	SweCens	109621554	109621556	Census	Husband
2	SweCens	109621556	109621554	Census	Wife

**Table CONTEXT**

Id	Id_D	Id_C	Type	Value	Timestamp
1	SweCens	109621555	Level	Household	1900-12-31
2	SweCens	109621555	Residence name	Örträsk	1900-12-31
4	SweCens	109621555	Household number	161	1900-12-31

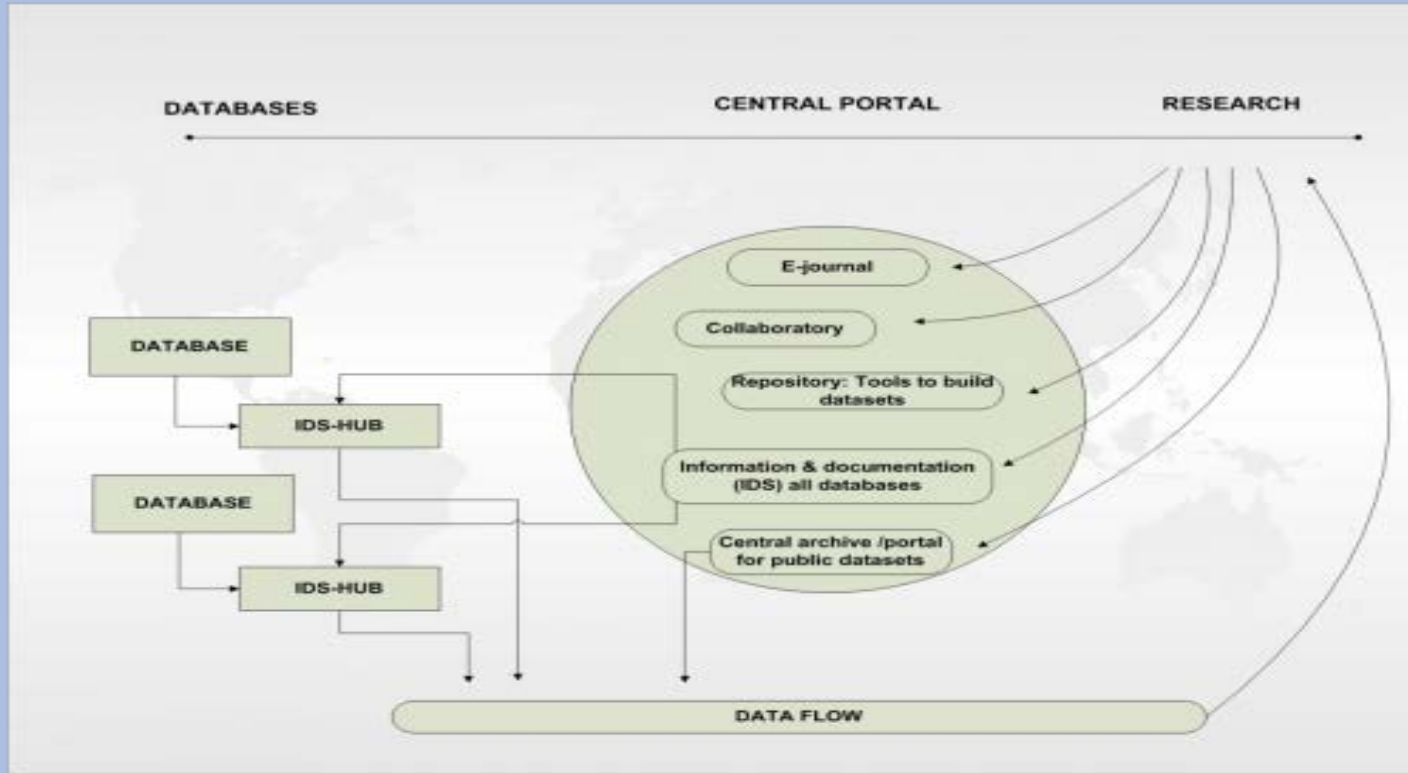
**Table INDIV\_CONTEXT**

Id	Id_D	Id_I	Id_C	Source	Relation
1	SweCens	109621556	109621555	Census	Household member



# European Historical Population Samples Network

www.ehps-net.eu



Tack för visat intresse!  
[mats.berggren@riksarkivet.se](mailto:mats.berggren@riksarkivet.se)