



Vetenskapsrådet

SWEDISH RESEARCH COUNCIL

The Register Utilisation Tool

A Practical implementation of GSIM as support in register-based research

Magnus Eriksson,

The Unit for Register-Based Research

Swedish Research Council

Register based research

- Sweden has many government registers, biobanks, health registers containing population based data.
- Data has been collected for decades and the registers are a valuable research resource.
- Time spent to grasp the meaning of the register variables and to communicate with the register owners before disclosure can sometimes add up to two years.



Swedish research Council Government Commission

Improve the conditions to use register data in research by:

- I) Establishment of a Register Data Board
- II) Provision of information, advisory and education functions for register-based research
- III) Improved access to registers for research purposes via the coordination of data delivery



Challenges for the researcher

Before starting a register based research project you might need an answer to some questions:

- Can register data be used to answer my research question?
- What data are there?
- What registers holds the appropriate data?
- What do the register variables mean?
- Are the meaning similar enough between registers?
- Can data from the registers be linked?

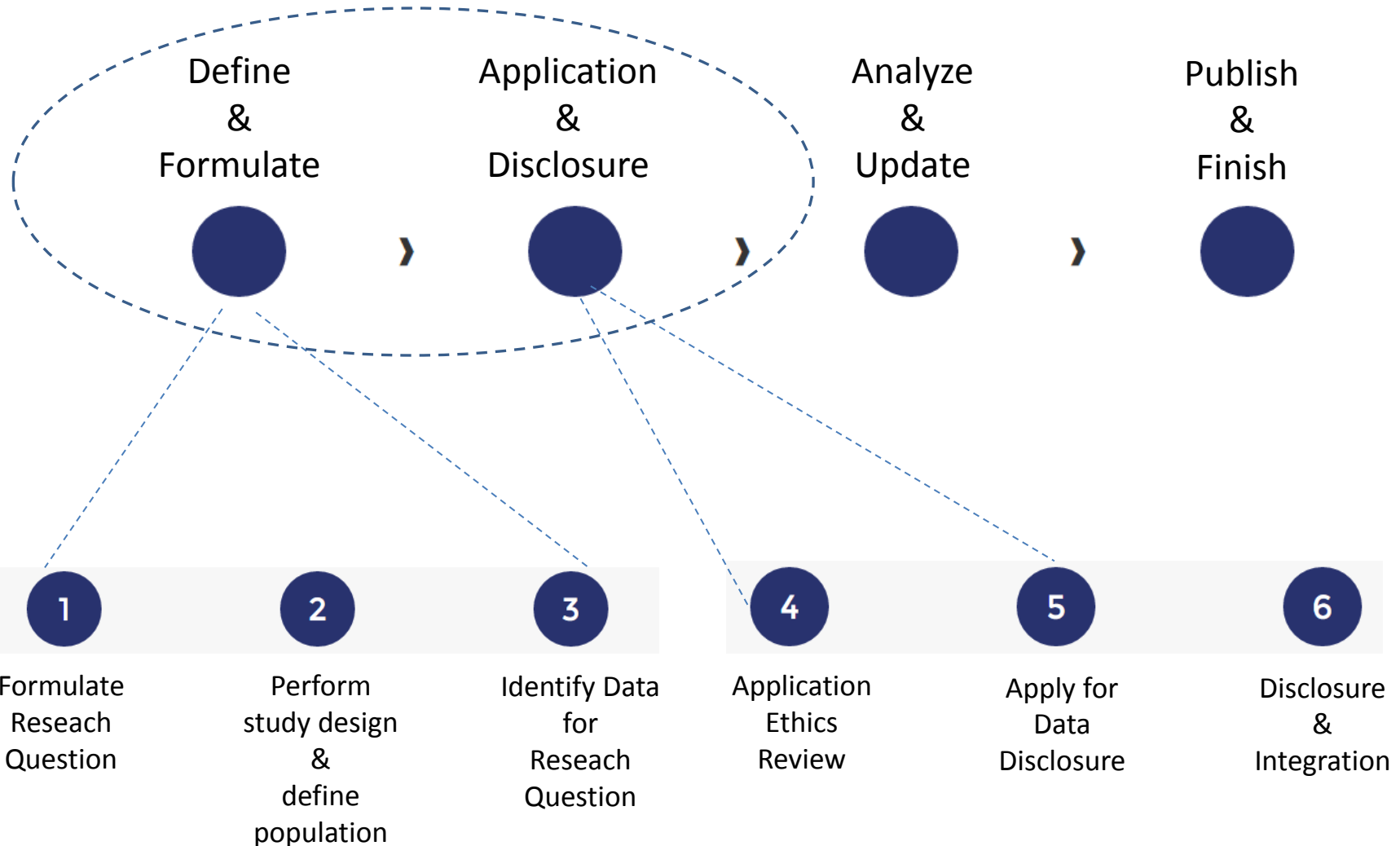


Register Utiliser Tool

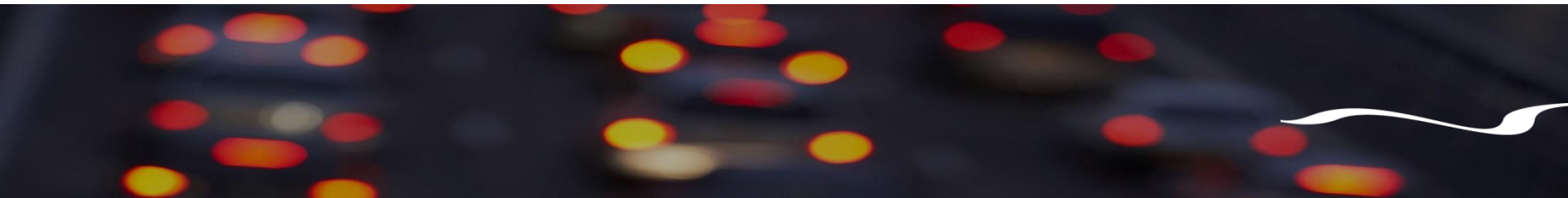
- Decreases the researchers “Time-To-Data”
 - time spent to grasp the meaning and representation of the register variables
 - time spent evaluating variables
 - time spent to communicate with the register owners before disclosure



RUT in the register-based research process

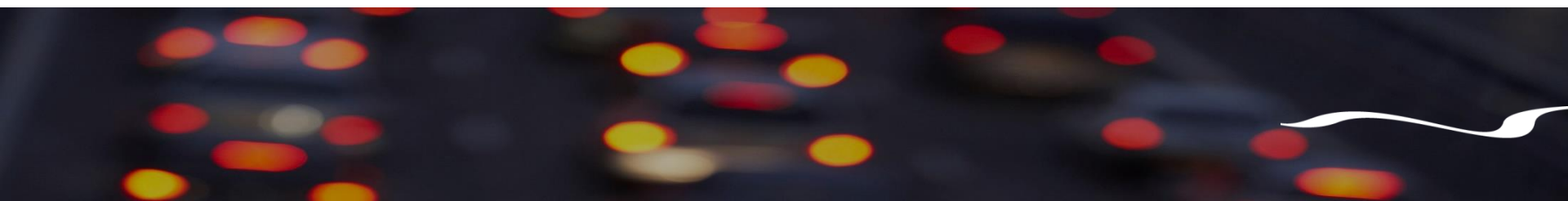


REQUIREMENTS



Prerequisite

The separation of metadata and data because of legal constraints.



High-level Functional Requirements

Support the researcher in selecting a study population.

Provide functions for variable search that do not require knowledge of register owners, registers or variable names and a way to search by meaning/concepts

Provide metadata to support the researcher during evaluation of a variable in relation to the research question.

Support the researcher during harmonization efforts by providing the metadata needed during harmonization analysis.

Give the researcher access to metadata on variable meaning, representation and populations in order to support communication of design in a clear and distinct way.

Provide support during analysis of a variables quality & sources, collection methods etc.

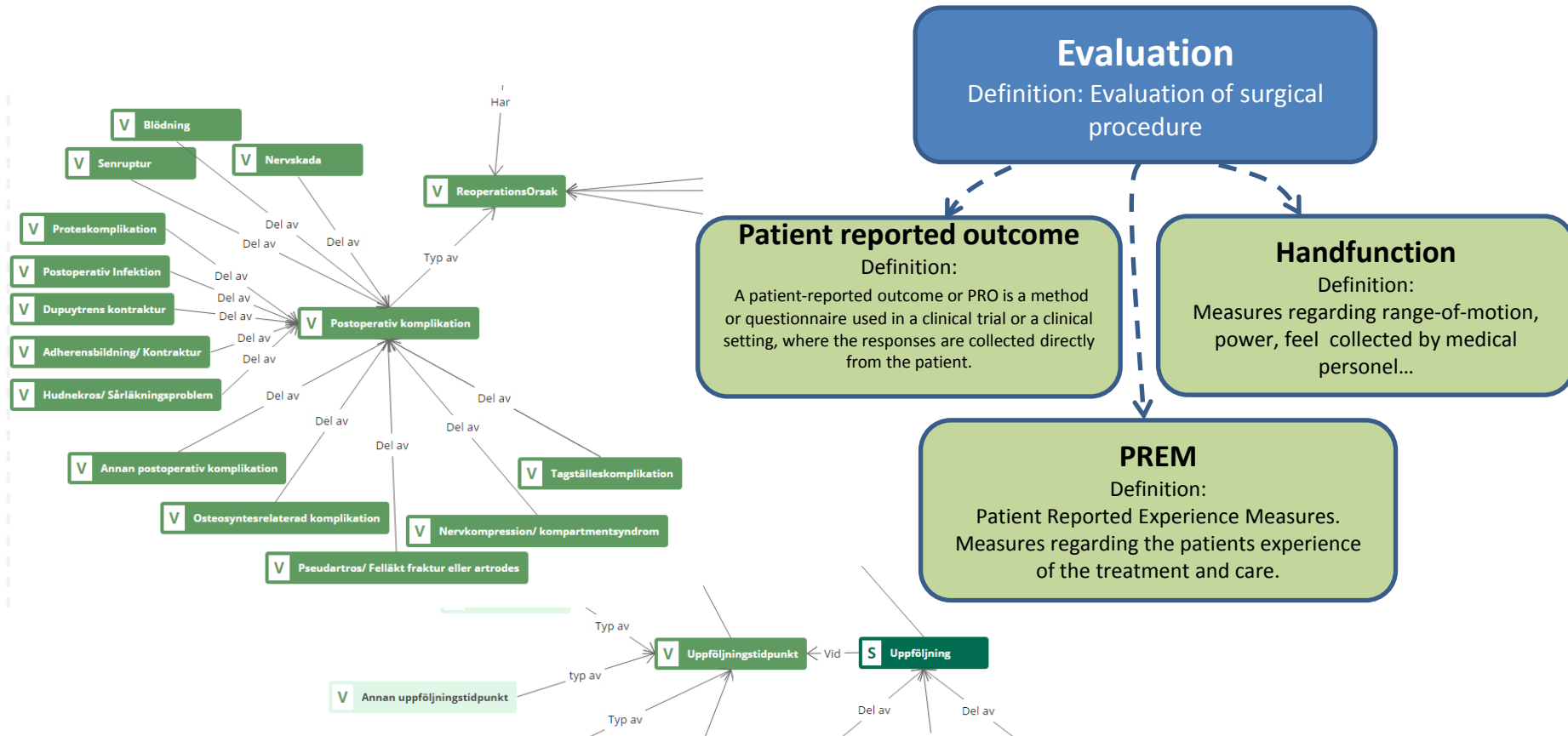
Search

- Without previous knowledge

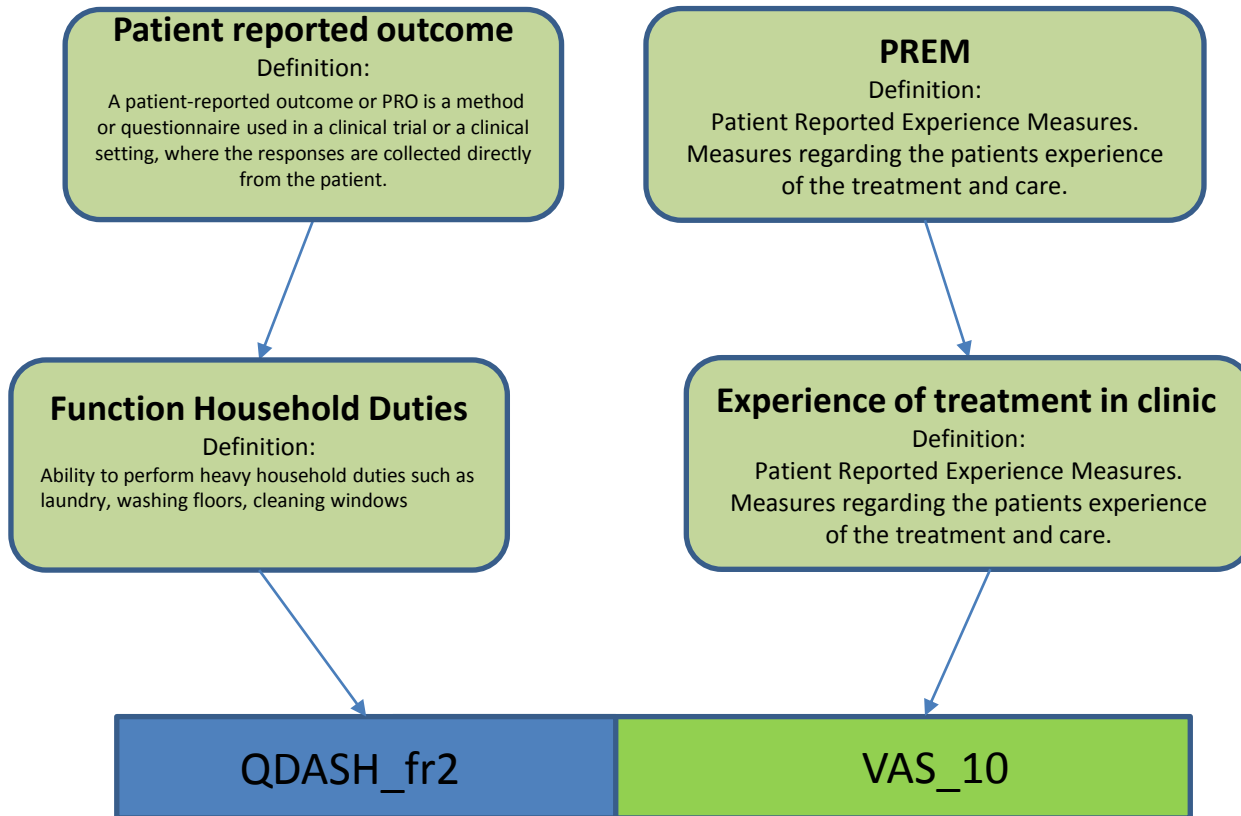


Search

- Through Concepts/ConceptSystems

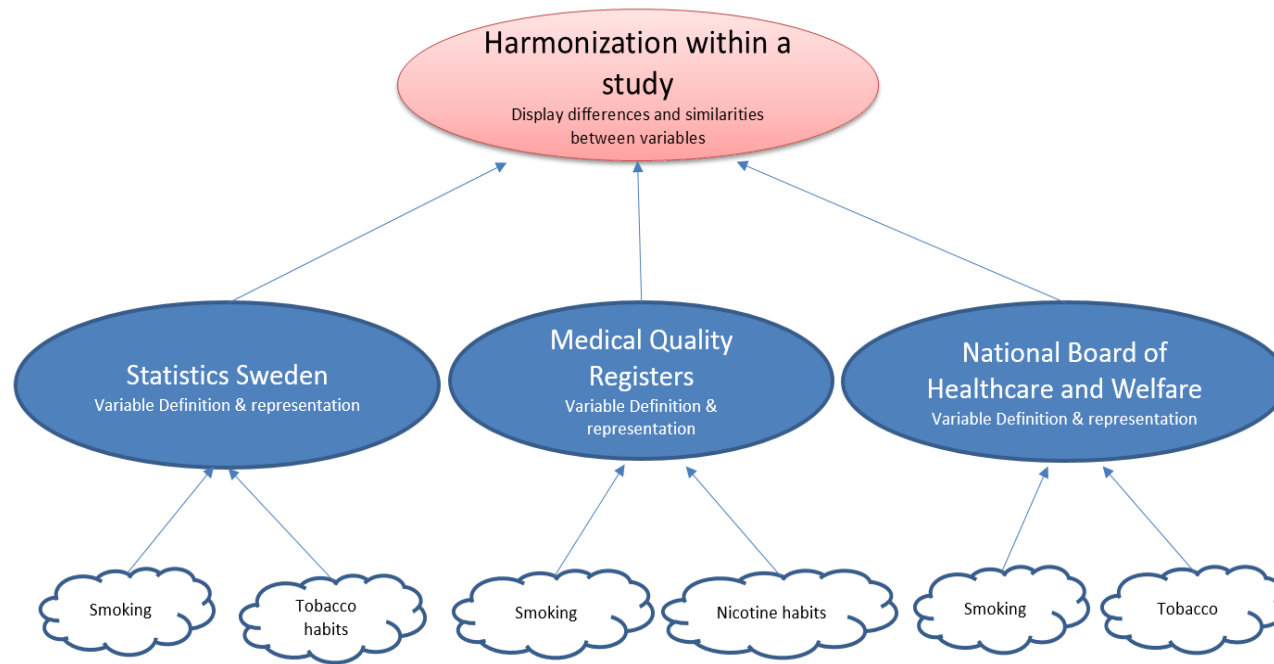


...to variables



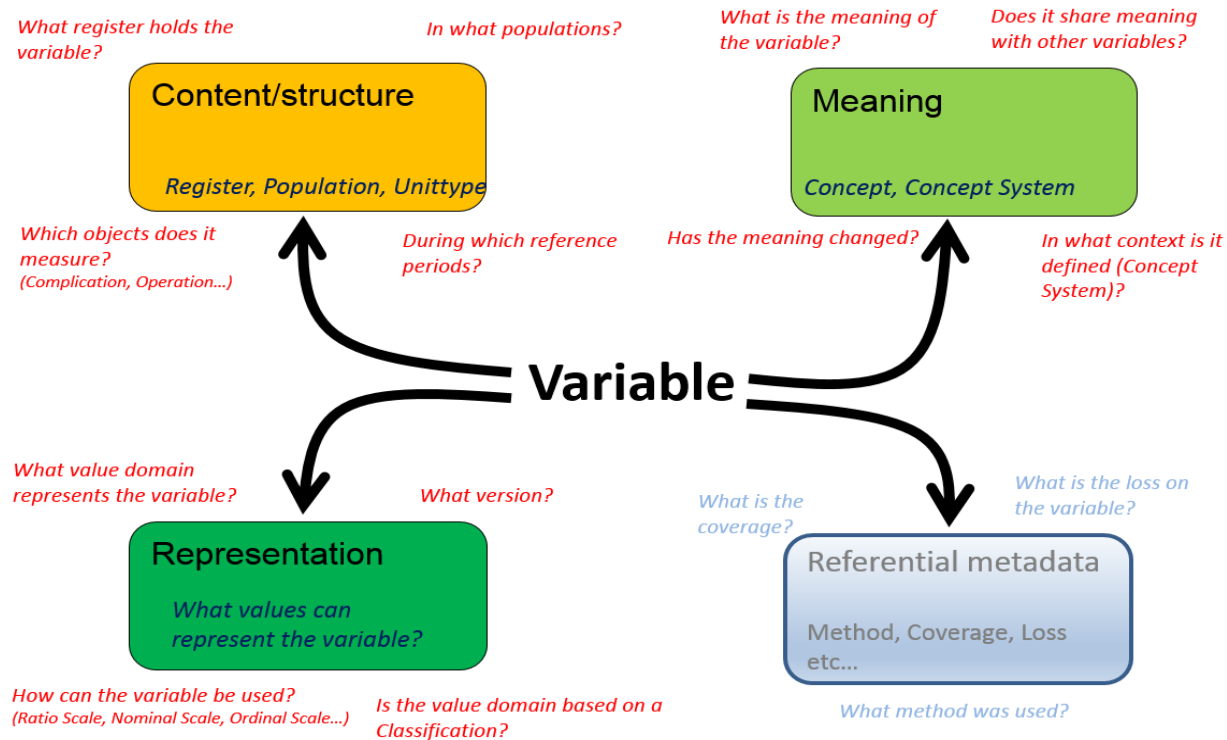
Harmonization Efforts

- Display differences & Similarities



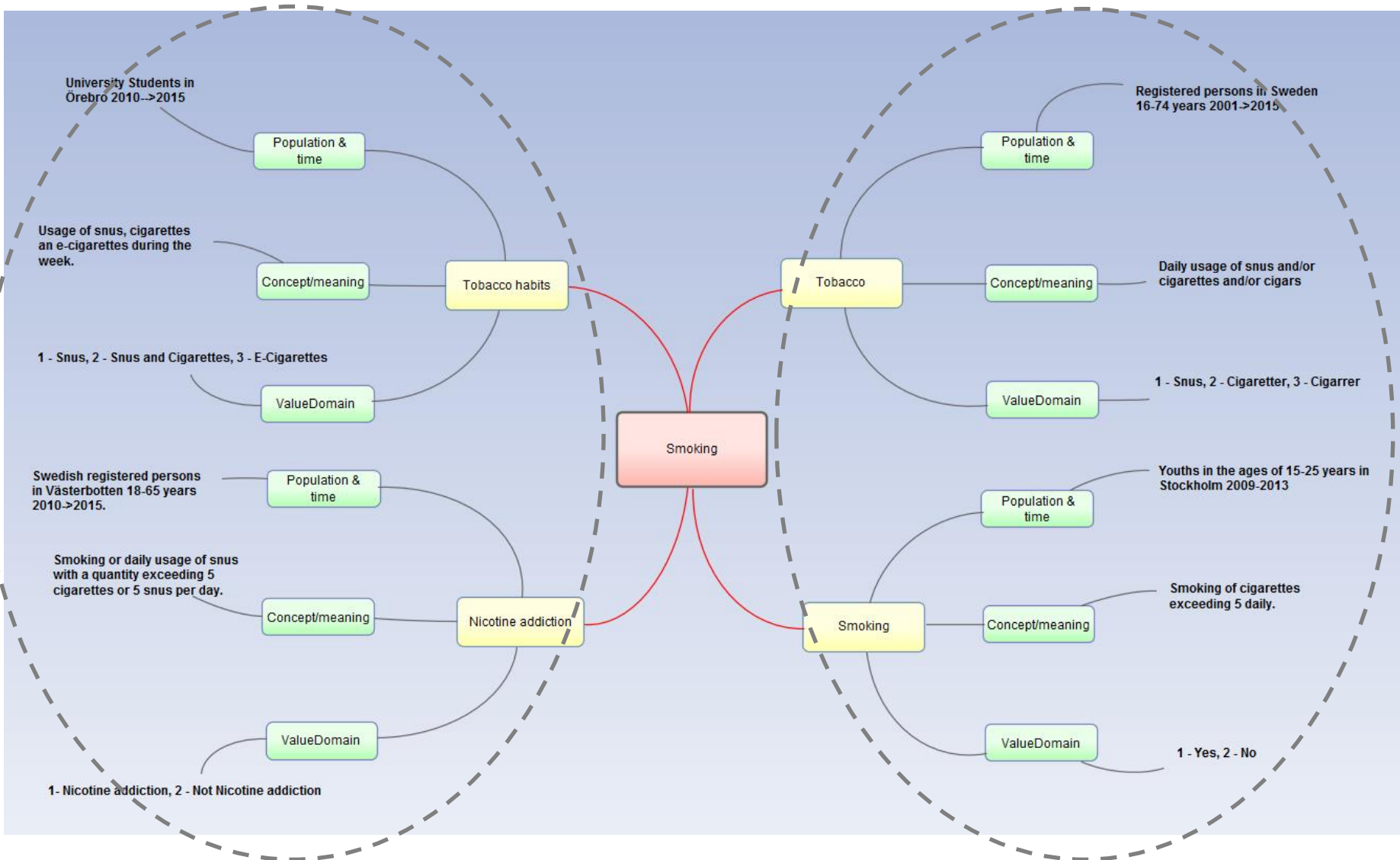
Harmonization Efforts

– Perspectives on Differences & Similarities



Harmonization Efforts

– Perspectives on Differences & Similarities

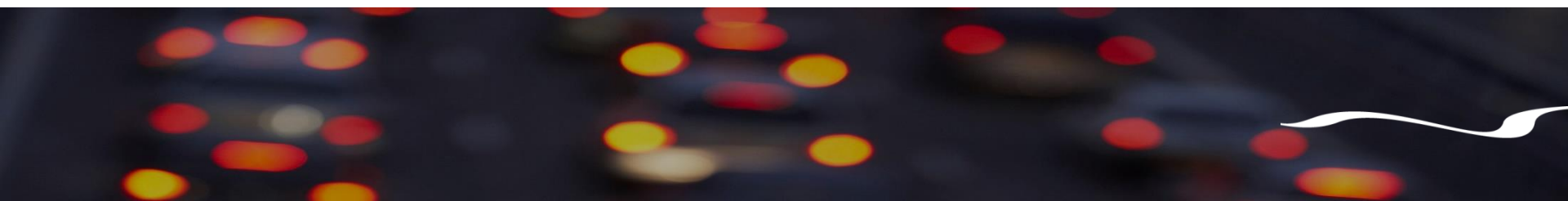
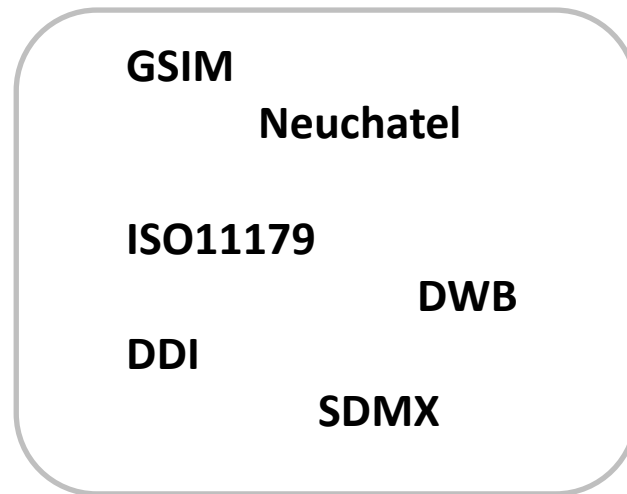


EVALUATION



Evaluation

- in relation to the project requirements



SELECTION



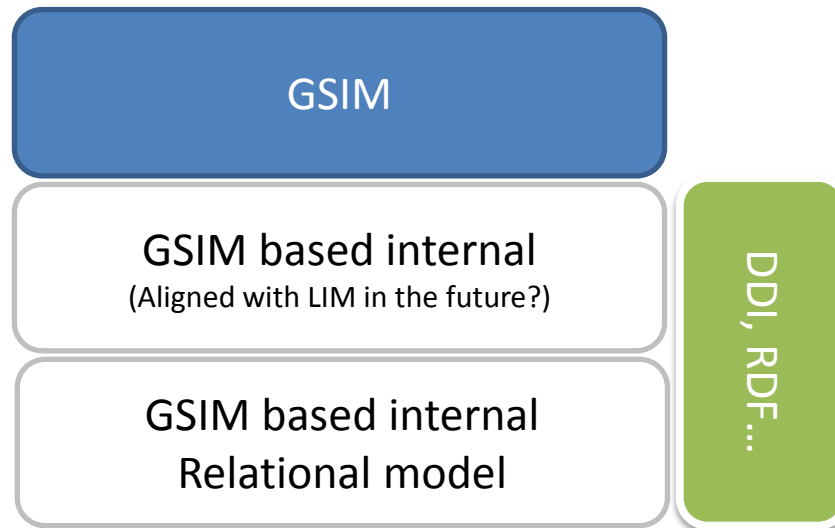
Conceptual standardized framework

-The choice of GSIM

- Separation of meaning from representation.
 - Provides a foundation for implementing search by variable meaning .
 - Enables the conceptual support that the researchers need during evaluation of a variable and harmonization efforts (same meaning different names and vice versa).
- Enabling comparison of variables and parts of variables.
 - Comparing parts of a variable during evaluation and harmonisation efforts.
- Strong support for administration of valuedomains/codelists/classification.
 - During evaluation of a variable and harmonisation efforts.
- Large degree of Business Domain independence
 - The register holders that provide the metadata come from a wide variety of business domains.
- Common language when communicating researchers metadata needs with register owners.
 - Model based and on conceptual level.
- Providing a structure to incorporate referential metadata.
 - Quality, sources and collection methods etc. For Collection Cycles, Populations, Variables etc.

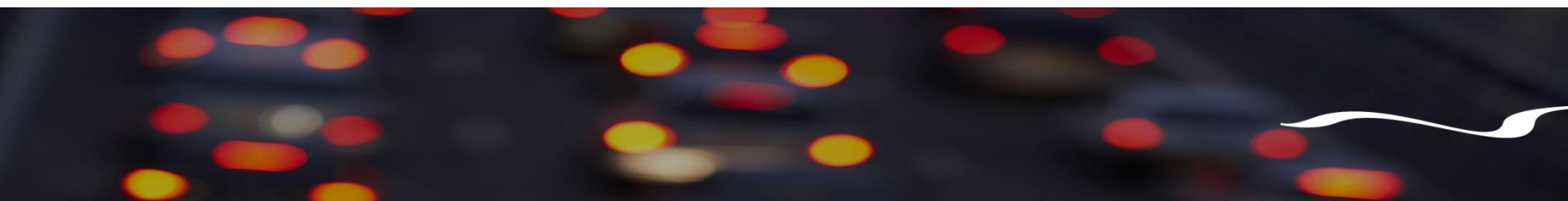


Logical, Physical & Exchange

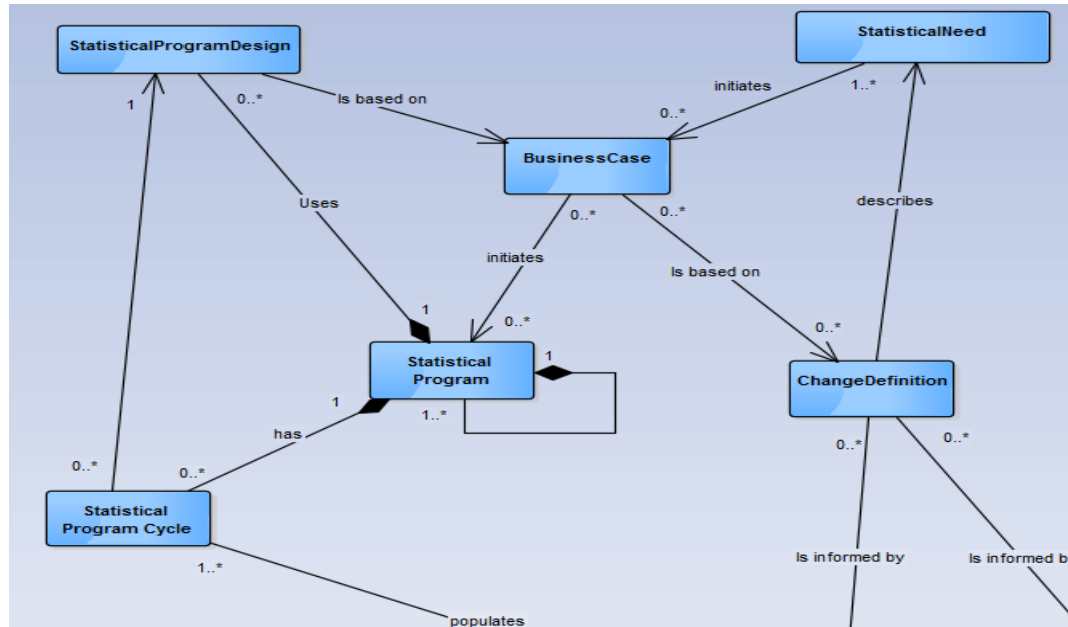


GSIM Selection

Mainly GSIM Concepts Group, some addition from the Business Group.

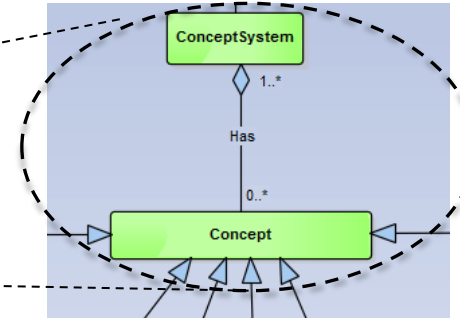
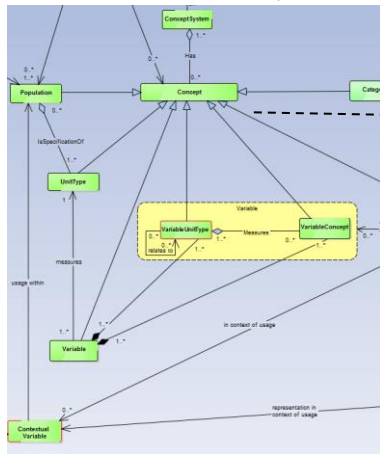


Objects regarding register, variants and change



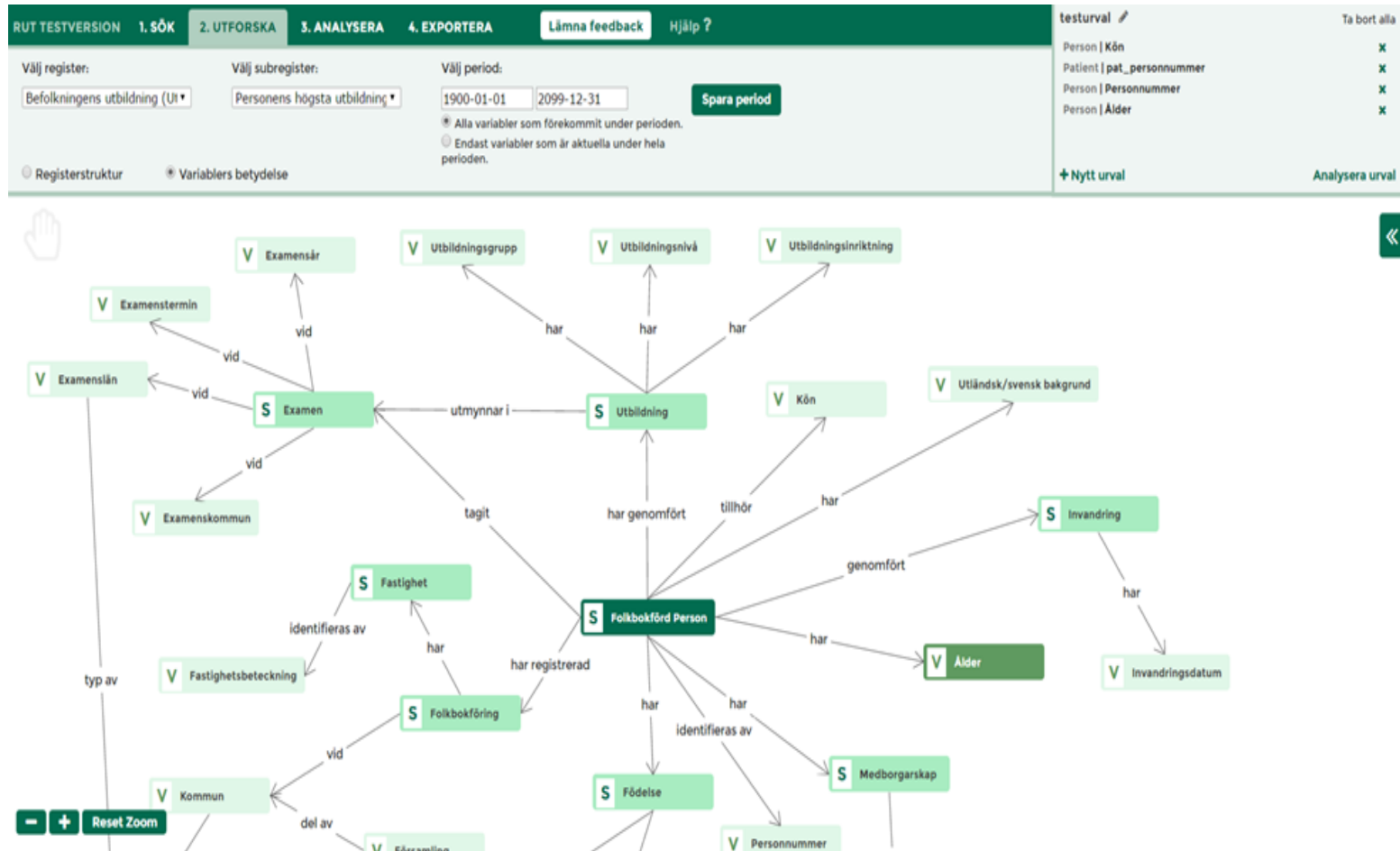
We added an association between “Statistical Program Cycle” and “Population” in order to express our usage of the Information Objects more clearly.

Concept System

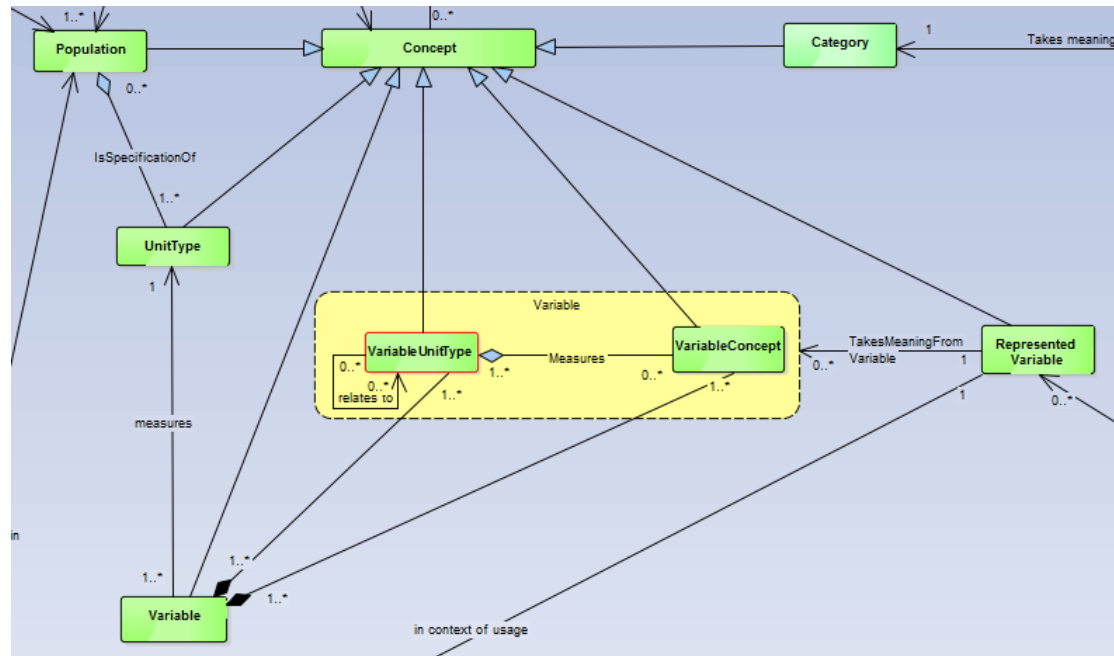


Gives the register holder an opportunity to present different perspectives of its Concepts in order to give the researcher an understanding of the intended meaning. Supports search and harmonisation efforts.

Presenting a Concept System in the application



Concept specializations (Population, Unit Type, Variable)



Variable in detail. We wanted to provide more detail to the GSIM variable to be able to visualize the variable as two parts, the "Variable UnitType" and the "Variable Concept", e.g. Father [Variable UnitType] + Income [Variable Concept].

Presenting variables in the application

The screenshot displays a web application interface for data exploration. At the top, there are navigation tabs: RUT TESTVERSION, 1. SÖK, 2. UTFORSKA (active), 3. ANALYSERA, 4. EXPORTERA, Ge återkoppling, and Hjälp ?. Below these are input fields for 'Välj register:' (Registret över totalbefolkning), 'Välj subregister:' (Födda), and 'Välj period:' (1900-01-01 to 2099-12-31). A 'Spara period' button is present. Radio buttons allow switching between 'Registerstruktur' and 'Variablers betydelse'. A '+ Skapa nytt urval' button is also visible.

On the right, a dropdown menu shows a list of population selections, with 'FÖDDA PERSONER I SVERIGE 2014-01-01' highlighted. A red dashed circle highlights this menu and a red arrow points to the '+ Skapa nytt urval' button.

The main area features a logical grouping diagram with three main nodes: 'Mor', 'Person *', and 'Folkbokföring'. 'Mor' and 'Person *' are connected by a 'har' relationship. 'Person *' is connected to 'Folkbokföring' by a 'har' relationship. 'Folkbokföring' is connected to 'Födelseland mor' by a 'har' relationship. 'Födelseland mor' is also connected to 'Person *' by a 'har' relationship. 'Mor' and 'Person *' are also connected to 'Födelseland mor' by 'del av' relationships.

The 'Födelseland mor' variable details are shown on the right, including its definition: "Födelse inom suverän stat. Även kallad självständig stat, självständigt land, oavhängigt stat eller eget land, är en stat, en entitet med ett definierat territorium där man utövar intern och extern suveränitet, där man enligt Montevideokonventionen har en permanent befolkning, en regering, självständighet från andra stater och makter, samt möjligheten att ingå relationer med andra suveräna stater.[1][2] Den suveräna staten är ofta underförstådd som en stat som inte är beroende av, eller under förutsättning av, någon annan makt eller stat.[3][4][5] Medan en suverän stat i abstrakta termer kan existera utan att erkännas av andra suveräna stater, är det ofta svårt för icke erkända stater att utöva full traktatbildande makt och engagera sig i diplomatiska relationer med andra suveräna stater."

The 'Värdeområde' section is highlighted with a red dashed circle and lists: "Länder och världsdelar enligt RTB", "Förekommande länder i RTB för födelseland respektive medborgarskapsland", "Domäntyp: Lista", "Värde: Nominellt värde", and "Giltighetstid: 1998-12-31 - 2014-12-31". A 'Lägg i urval' button is at the bottom right.

We then get the opportunity to provide a visualization of the logical grouping of Variable Concepts within Variable UnitTypes.

Evaluating & selection variables

(Adding representation & time)

Historik Värde mängd

Namn	Beskrivning	Domäntyp	Värdetyp	Giltig fr.o.m	Giltig t.o.m
SUN 2000 - Utbildningsinriktning	SUN 2000 är en standard för klassificering av enskilda utbildningar. Inriktningsmodulen är hierarkiskt uppbyggd där den fjärde siffran i inriktningsmodulen utelämnas om de tre första ger tillräcklig information om utbildningens inriktning	Lista	Ordinalvärde	2000-01-10	2013-06-15
SUN enligt MIS 1996:1	Svensk UtbildningsNomenklatur, SUN, enligt den äldre systematik som gällde fram till år 2000 (MIS 1996:1).	Lista	Ordinalvärde	1997-01-10	1999-06-15
Utbildningens inriktning enligt MIS 1996:1	Svensk UtbildningsNomenklatur, SUN, enligt den äldre systematik som gällde fram till år 2000 (MIS 1996:1). Den första positionen av den 5-ställiga SUN-koden angav utbildningens inriktning.	Lista	Ordinalvärde	1997-01-10	1999-06-15

BESKRIVNING VÄLJ TILL EXPORTLISTAN

Till Export	Subregister	Insamlingscykel	Population	Beskrivning	Definition	Referensperiod fr.o.m.	Referensperiod t.o.m.	Geografi
<input type="checkbox"/>	Personens högsta utbildning,16-74 år	Högsta utbildning t.o.m. värterminen 1997 hos den folkbokförda befolkningen 1997-12-31 i åldern 16-74 år.	Folkbokförda personer i Sverige 16-74 år 1997-12-31	Befolkningsregistrets population i åldern 16-74 år vid årets slut	Folkbokförda personer i Sverige 16-74 år	1997-12-31	1997-12-31	Sverige
<input checked="" type="checkbox"/>	Personens högsta utbildning,16-74 år	Högsta utbildning t.o.m. värterminen 1997 hos den folkbokförda befolkningen 1997-12-31 i åldern 16-74 år.	Folkbokförda personer i Sverige 16-74 år 1997-12-31	Befolkningsregistrets population i åldern 16-74 år vid årets slut	Folkbokförda personer i Sverige 16-74 år	1997-12-31	1997-12-31	Sverige
<input checked="" type="checkbox"/>	Personens högsta utbildning,16-74 år	Högsta utbildning t.o.m. värterminen 1998 hos den folkbokförda befolkningen 1998-12-31 i åldern 16-74 år.	Folkbokförda personer i Sverige 16-74 år 1998-12-31	Befolkningsregistrets population i åldern 16-74 år vid årets slut	Folkbokförda personer i Sverige 16-74 år	1998-12-31	1998-12-31	Sverige
<input type="checkbox"/>	Personens högsta utbildning,16-74 år	Högsta utbildning t.o.m. värterminen 1998 hos den folkbokförda befolkningen	Folkbokförda personer i Sverige 16-74 år	Befolkningsregistrets population i åldern 16-74 år	Folkbokförda personer i Sverige 16-74 år	1998-12-31	1998-12-31	Sverige

RUT TESTVERSION 1. SÖK 2. UTFORSKA 3. ANALYSERA 4. EXPORTERA Ge återkoppling Hjälp ?

Exempel

Exempel

Med markerade **Utför**

Mor | Ålder vid händelsen mor

→ ÖPPNA I UTFORSKA

Far | Civilstånd far

→ ÖPPNA I UTFORSKA

Utbildning | Utbildningsinriktning

→ ÖPPNA I UTFORSKA

BESKRIVNING VÄLJ TILL EXPORTLISTAN

Registerhållare
STATISTISKA CENTRALBYRÅN

Register
Befolkningens utbildning (Utbildningsregistret)

Subregister
Personens högsta utbildning,16-74 år

Beskrivning

Definition

De flesta utbildningar består av flera olika ämnen/ämnesinriktningar. Vid klassificering av utbildningens inriktning är det utbildningens huvudinnehåll som bestämmer vilken kod som en viss utbildning skall ha.

Värde mängd: SUN 2000 - Utbildningsinriktning

SUN 2000 är en standard för klassificering av enskilda utbildningar. Inriktningsmodulen är hierarkiskt uppbyggd där den fjärde siffran i inriktningsmodulen utelämnas om de tre första ger tillräcklig information om utbildningens inriktning

Domäntyp: Lista

Värdetyp: Ordinalvärde

Giltighetstid: 2000-01-10 - 2013-06-15

Insamlingsmetod: Longitudinal

REGISTERFORSKNING.SE

