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 Comission

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

Define & Formulate

Application & Disclosure

Analyze & Update

Publish & Finish

1. Formulate Research Question
2. Perform study design & define population
3. Identify Data for Research Question
4. Application Ethics Review
5. Apply for Data Disclosure
6. Disclosure & Integration
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

- Want to search without prior knowledge of register owners or registers!
- Want to search on meaning not only on (obscure) variable names!
- Want to be able to restrict search by time and population!

Search:
- QDash
- No of Cigaretts
- Nicotine addiction
- VAS
- Smoking
- Tobacco
- Tobacco habits
Handfunction

Definition:
Measures regarding range-of-motion, power, feel collected by medical personnel...

Patient reported outcome

Definition:
A patient-reported outcome or PRO is a method or questionnaire used in a clinical trial or a clinical setting, where the responses are collected directly from the patient.

PREM

Definition:
Patient Reported Experience Measures. Measures regarding the patient's experience of the treatment and care.
...to variables

**Patient reported outcome**
Definition:
A patient-reported outcome or PRO is a method or questionnaire used in a clinical trial or a clinical setting, where the responses are collected directly from the patient.

**Function Household Duties**
Definition:
Ability to perform heavy household duties such as laundry, washing floors, cleaning windows.

**PREM**
Definition:
Patient Reported Experience Measures. Measures regarding the patients experience of the treatment and care.

**Experience of treatment in clinic**
Definition:
Patient Reported Experience Measures. Measures regarding the patients experience of the treatment and care.

**QDASH_fr2**

**VAS_10**
Harmonization Efforts
- Display differences & Similarities

Harmonization within a study
Display differences and similarities between variables

Statistics Sweden
Variable Definition & representation

Medical Quality Registers
Variable Definition & representation

National Board of Healthcare and Welfare
Variable Definition & representation

Smoking
Tobacco habits
Smoking
Nicotine habits
Smoking
Tobacco
Harmonization Efforts
– Perspectives on Differences & Similarities

Variable

Content/structure
Register, Population, Unittype
- What register holds the variable?
- In what populations?
- Which objects does it measure? (Complication, Operation...)
- During which reference periods?

Meaning
Concept, Concept System
- What is the meaning of the variable?
- Does it share meaning with other variables?
- Has the meaning changed?
- In what context is it defined (Concept System)?

Representation
- What values can represent the variable?
- What version?
- What value domain represents the variable?
- Is the value domain based on a Classification?
- How can the variable be used? (Ratio Scale, Nominal Scale, Ordinal Scale...)

Referential metadata
- Method, Coverage, Loss etc...
- What method was used?
- What is the coverage?
Harmonization Efforts – Perspectives on Differences & Similarities
Evaluation
- in relation to the project requirements

GSIM
Neuchatel
ISO11179
DWB
DDI
SDMX
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

GSIM based internal
(Aligned with LIM in the future?)

GSIM based internal
Relational model

DDI, RDF...
GSIM Selection

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