|  |  |
| --- | --- |
| PCA/MFI  Klas Blomqvist  klas.blomqvist@scb.se |  |

# HLG Project proposal, GSIM-variables

## Background

The Generic Statistical Information Model (GSIM) contains a number of groups. Of these, the Concepts Group is the most established and elaborated. The main reason for this is because the Concepts Group is the area influenced by traditional statistical metadata development in statistical organizations.

These organizations have developed a lot of experience over the years, and several organizations are now using GSIM in development projects. Earlier experiences with standards such as ISO 11179 show that their implementation in statistical organisations can be quite different due to differences in interpretation.

There is a risk that the same thing will happen to GSIM unless there is a common understanding on how to interpret and use the model.

Discussions regarding GSIM variables and related concepts (concept systems, concept, variable, represented variable and instance variable) are showing that they can all be interpreted in different ways in different GSIM implementations: If this continues, the potential gain from CSPA and other projects on tools and methods will suffer, as issues are solved in different ways in each implementation.

## Project objectives

The main objectives of this project are to:

* Create a GSIM annex on variables, similar to the existing one on classifications
* Provide a clearer understanding of how to interpret and implement GSIM concepts, focusing on the variable parts and their implementation across the GSBPM (i.e. what are the roles that variables can play in the data structures produced in each of the GSBPM phases)
* Reduce the risk of high costs and comprehensive hands on work in order to implement CSPA components developed by other organizations due to different interpretations of GSIM concepts
* Map the GSIM concepts with the corresponding ones in DDI and SDMX

## Scope

The scope of the project would be to clarify the use of variables and their roles within GSIM, based upon the experiences already gained from GSIM implementations.

Such work would help to provide a solid foundation for the use of GSIM to facilitate collaborations using CSPA and within other modernization/standardization initiatives.

Using examples from previous GSIM implementations would provide a clear framework that maximises interoperability between statistical organizations.

Providing further clarification of what a variable is would involve specifying:

* + - * The possible roles of variables in a statistical process;
* How concept systems should be created and handled in GSIM
* Roles for concept management in GSIM

Organizations often view these questions differently, so it also involves consideration of concept management, of how to define concepts and concept systems.

The need is growing for a high level of consistency in the interpretation of variables and concepts, to allow comparison and automatic linking between statistical organizations. This is because many organizations are using DDI and SDMX for storing and managing concepts and variables.

The lack of a clear and commonly shared interpretation of concepts and variables jeopardizes the use of DDI and SDMX as tools for an easy exchange of data and metadata between different organizations. Hence, mapping to DDI or SDMX is an important deliverable of this work

This project also has links to the work on “DDI moving forward”.

## Contents

**Work Package 1: GSIM-Variables annex**

This work would develop a detailed reference guide/model for variables (similar to the approach adopted for the GSIM Statistical Classification Model).

It would be based on input from experts who have experience of implementing GSIM in their organizations.

Activities:

1. Reviewing of the following GSIM information objects related to variables

* Concepts
* Variables
* Represented variables
* Instance variables

1. Analysing case examples of the different implementations of GSIM.
2. Reviewing existing metadata standards such as the Neuchâtel Terminology Model for Variables.

Output:

A GSIM annex containing a reference guide /model that can be made available alongside GSIM. The annex shall contain a set of recommendations of how to interpret relevant objects in the GSIM Concepts group. The guide should include variables as inputs and as outputs across the GSBPM processes, and would include links to other GSIM objects.

**Work Package 2: Best practices in managing variables in a statistical system**

This would include best practice guidelines on how to implement GSIM, i.e. how to organize variables in a data structure, specifying the different roles these variables can play in micro or macro data sets, throughout the whole production process.

Activities:

1. Creating examples of use cases of:
   1. How variables can contribute to the definition of a data structure and how they relate to the variables detected in previous GSBPM phases;
   2. Other uses (i.e., “exchange” variable model, CSPA, LIM)
2. Developing a Best Practices document as part of the GSIM-Variables annex.

Output:

* Detailed use cases containing examples of how variables flow across the phases in GSBPM and how they characterize data structures in each GSBPM phase.

General descriptions of the use cases.

**Work Package 3: Enhancing Interoperability**

To describe how different standards (such as GSIM, DDI and SDMX) are related.

Activity:

* Mapping the GSIM variables objects to DDI, SDMX, and monitoring the existing initiatives of mapping GSIM to DDI and SDMX.

Output:

* Detailed mapping schemes that display the relations between the standards.

## Definition of success

The project will be successful if it results in an improved understanding, and a common way of interpreting, the variables part of GSIM-Concepts. This will in turn lead to easier to implementation of CSPA tools in statistical organizations.

The project will produce a GSIM appendix with elaborated examples on how to interpret GSIM-variables. Similar work on the appendix for GSIM classifications has demonstrated the value of such material, and maximises the potential use of GSIM in statistical organizations.

This project is not about content-focused implementations, but rather for developing tools to assist statistical organizations in their work.

## Expected costs

| **Work Package** | **Estimated resources** | **Source of resources** | **Other costs (in US Dollars)** |
| --- | --- | --- | --- |
| GSIM Variables annex | 6 person months | Volunteer NSOs plus UNECE Secretariat |  |
| Best practices in managing variables in a statistical system | 4 person months | Volunteer NSOs plus UNECE Secretariat |  |
| Enhancing Interoperability | 4 person months | Volunteer NSOs plus UNECE Secretariat |  |
| Project management | 6 person months | A project manager working in the UNECE Secretariat. Input from Executive Board and HLG members (in their role as project sponsors) | Travel costs for project events |
| **Total** | 20 person months |  |  |

## Timetable

The project will aim to complete its activities by the end of 2016. There are, however, various unknowns which may affect the timetable:

* The availability of resources from national and international statistical organizations to support this project – if the necessary resources are not available, either the timetable will need to be extended, or the outputs will need to be re-defined (in terms of quality or quantity or both)
* The availability of project management and support resources. This will require either financial contributions to allow suitable persons to be hired, or donations in kind, such as the allocation of an expert to assist with project management. If this is not forthcoming, the outputs will need to be re-defined.

## Project governance

The project sponsor is the HLG, which is the group that has ultimate responsibility for signing off the project deliverables. In practice, this responsibility will be delegated to the Executive Board.

A project manager will have day-to-day responsibility for the running of the project, providing regular updates and signalling any issues to the Executive Board as necessary.

If this project does not go ahead there is a big risk that each use of GSIM, such as national tools and methods development or CSPA initiatives, will take a different approach. This will result in the need for hands on mapping work, which could be avoided if variables in GSIM are used more consistently.

**Work method**

* Online meetings, common workspace on the web
* Sprint in connection to the MC standards workshop?

**Competences**

* Information/enterprise architects
* Metadata experts
* Methodologists
* Project manager