

German Federal Statistical Office: use of GSIM

Case study: German Federal Statistical Office German Federal Statistical Office: Use of GSBPM German Federal Statistical Office: use of GSIM

Contact person*	Dr. Sven Claußen
Job title	Assistant Head of section "Cross-Cutting IT Processes Relating to Metadata and Data Quality"
Email	metadata@destatis.de
Telephone	

Statistical Information Model

The German Statistical Verbund didn't have agreed to a standardized information model in the past. Many production systems and tools are in use which were optimized to specific tasks and brought along their own information model. An overarching interoperability between different tools is not given unfortunately.

With the adoption of the GSBPM an internationally standardized process model is available to connect the existing tools and systems along the production chain structurally. Together with the national efforts to standardize the production processes and with the consistent support of sub-processes with optimized production tools the need for a standardized information model has arisen to exchange information objects along the production chain.

Since 2012 several metadata experts meet regularly to develop a unified information model (besides other metadata tasks) to prepare a proposal for a national standard within the German Statistical Verbund. To ensure the largest possible compatibility to the already existing production tools and systems their integrated information models should be taken account of.

Therefore in the beginning the already existing information models have been documented and analyzed regarding differences and similarities. Then a core model has been developed which uses GSIM as a reference model and which combines fundamental common features of the existing models. With the help of this core model it is already possible to map the most important information objects. The next steps in 2015/2016 will enhance the core model to a more detailed information model which facilitates the technical transformation of the objects and their reuse in different systems and tools. The compatibility to GSIM shall remain preferably strong and comprehensive.

The alignment to GSIM consists at the top level in the adoption of the classification of five groups. Within these groups several categories are arranged which contain the information objects:

- Base
 - Base Artefacts
 - Organization
- Business
 - Process Model
 - Tools
 - Statistical Activity
 - Information Objects
- Exchange
 - Instruments
- Concepts
 - Glossary
 - Variables
 - Classifications
- Structure
 - Data Structure
 - Quality Information

1. Base

The group „Base“ contains the category „Base Artefacts“ which encompasses basic properties for every information object of the core model, e.g. „Owner“ and „i18n_text“ to manage multilingual text (label, denotation, description etc.).

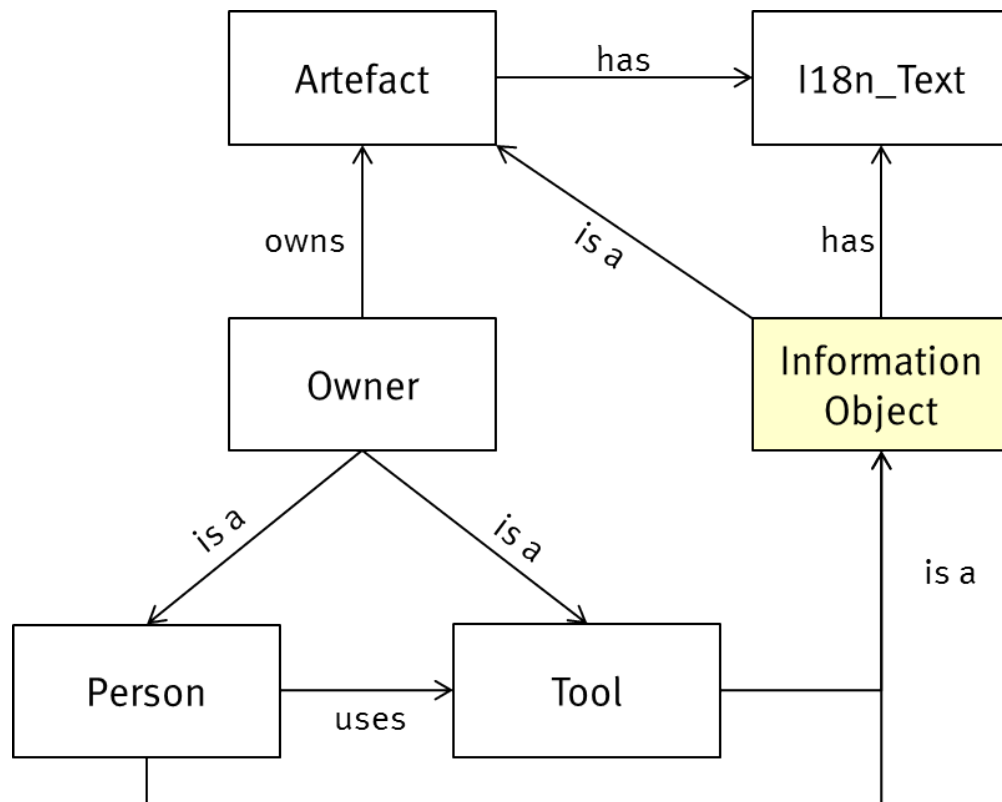


Figure 1: Sub-Model „Base Artefacts“

The information object „Artefact“ corresponds in general to the GSIM objects „Identifiable Artefact“ and „Administrative Details“. By using the attributes „version“ and „owner“ additionally Germanys core model conforms to SDMX as well.

The core model contains within the group “Base” the category “Organization” which encompasses information objects for holding administrative information about contacts and responsibilities. These objects are usually personalized. Persons are related closely to organization units. Both are connectable to many other information objects.

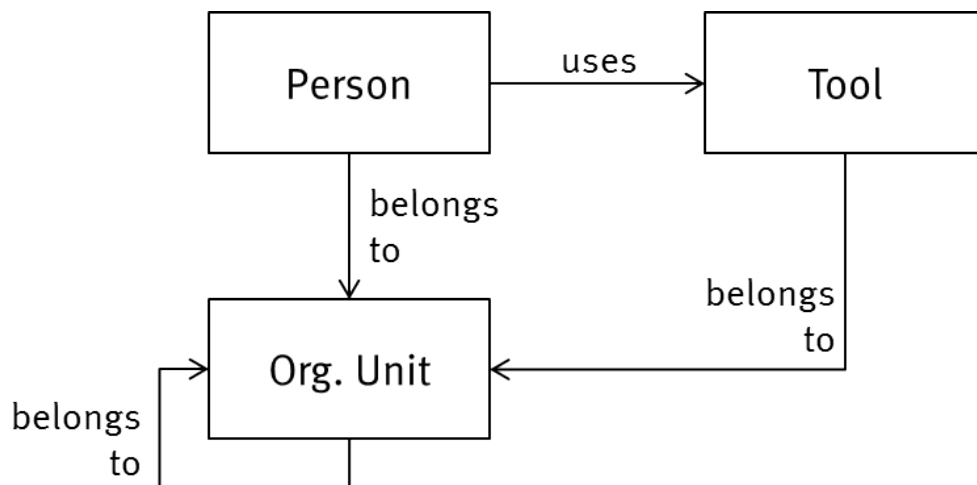


Figure 2: Sub-Model „Organization“

GSIM 1.1 arranges administrative information objects in the base group, too, and uses the information objects „Individual“ and „Organization Unit“ equivalently to „Person“ und „Organization Unit“. Both are children of the object „Agent“, which could be assigned any „Agent Role“. Specific „Contact Details“ cannot be found in GSIM 1.1.

2. Business

The object group „Business“ contains information objects which especially describes the design and the planning of the statistical program. The most important objects are „Statistical Need“, „Information Request“, „Business case“, „Statistical Program“, „Assessment“, the „Statistical Activity“ and objects for the data collection (like e.g. „Data Channel“, „Instrument“, „Question Block“, „Question“ etc.). The group is divided into four categories.

In order to map information objects to associated phases and sub-processes of the process model the information about processes are components of the core model, too. The process model can be described by only one information object („Process“). Sub-processes are processes with parents, phases are processes without parents. In addition „Process Model“ contains information about the process model itself (e.g. name, version).

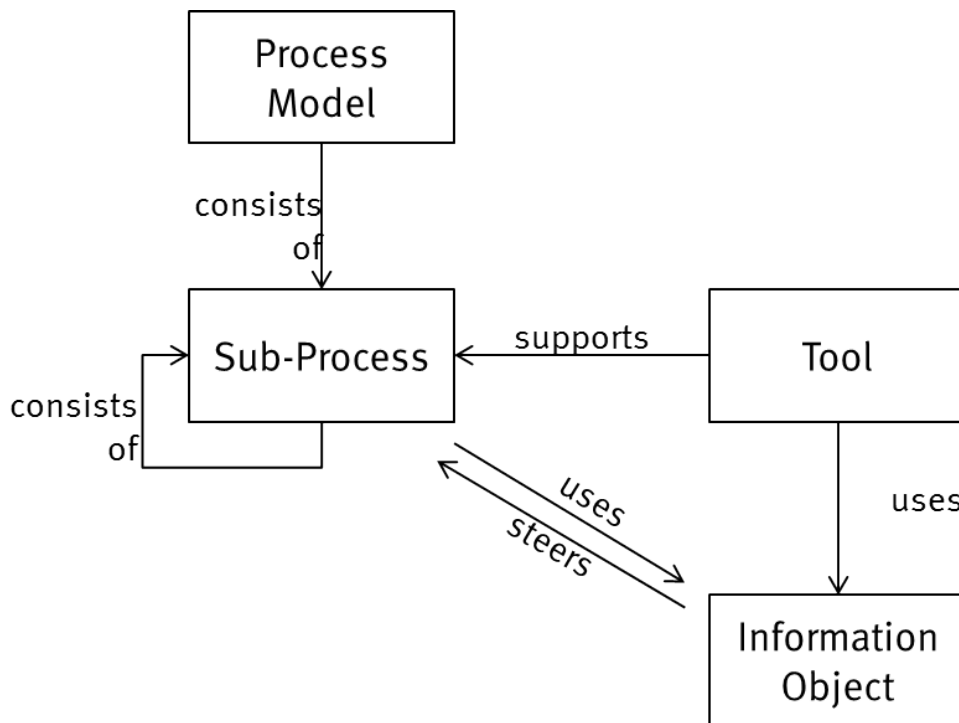


Figure 3: Sub-Model „Process Model“, „Tools“ and „Information Object“

As equivalent to the business processes of the GSBPM the „Business Functions“ of GSIM 1.1 take up the business prospect to the production processes. Their purpose is to connect contacts, processes and resources (e.g. tools). As opposed to this a „Process Step“ describes the specific production process in more detail for a certain production cycle and binds input and output objects as well. While a „Process Step“ is especially suitable for production control the „Business Function“ is used to describe the general purpose of a process step. The information object „Process Step Design“ is used by GSIM to connect a detailed process description („Process Method“) to a certain process step.

The information object „Tool“ documents the assigned production tools and systems. In GSIM 1.1 there seems to be no equivalent object.

One of the key components of the core model are statistics and the information objects „Statistical Program“, „Statistical Activity“ and „Statistical Program Cycle“ respectively.

The information objects „Data Collection Variable“, „Population“ and „Periodicity“ are attached to the „National Law“ which is a subtype of „Legislation“.

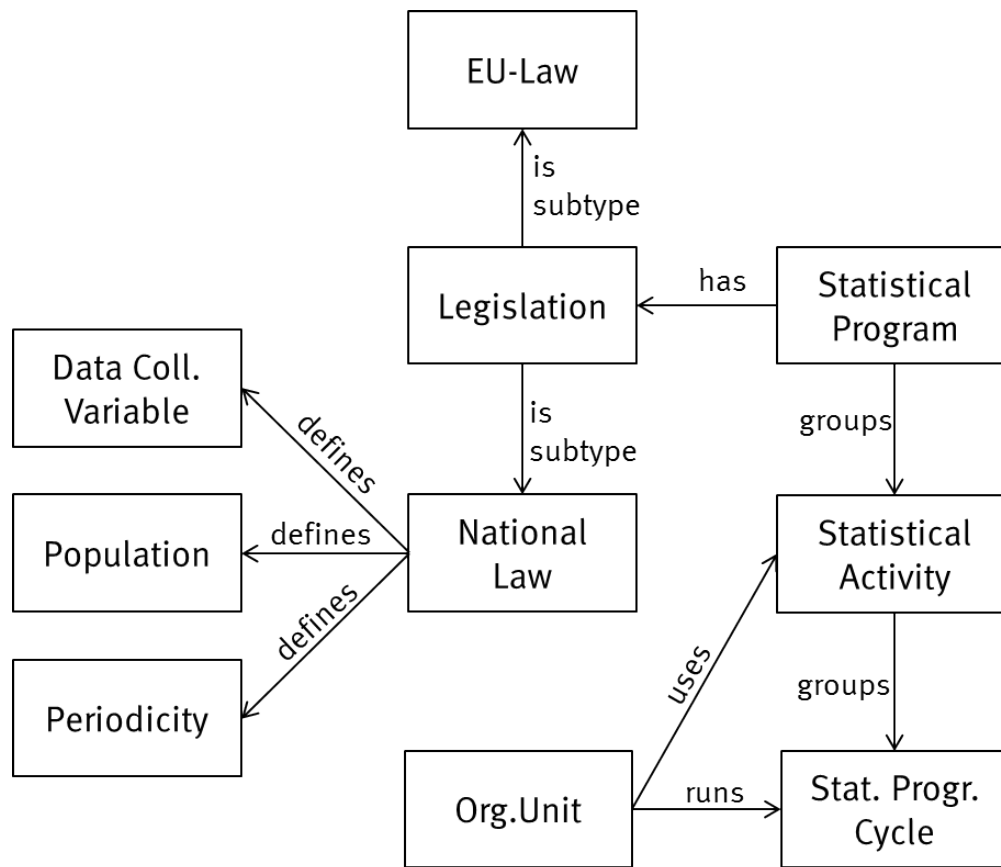


Figure 4: Sub-Model „Statistical Program“, „Statistical Activity“, „Statistical Program Cycle“ and „Legislation“

GSIM 1.1 structures the „Statistical Program“ within the Business group. A statistic in German linguistic usage is seen as „Statistical Program“ and every production cycle is described by „Statistical Program Cycle“ which is designed by the „Statistical Program Design“. At this point a strong connection to the process model exists, which is the ideal-typical template for a certain production cycle. A „Statistical Program Cycle“ consists of well-defined „Business Processes“, which uses „Business Services“ and is made up of „Process Steps“. „Process Steps“ are designed accordingly to the „Statistical Program Design“.

As with the process model and the associated processes information about the objects itself and their usage by the processes will be modeled by the information objects „Information Object“ and „Attribute“. „Mapping“ describes rules for the transformation of objects to another objects.

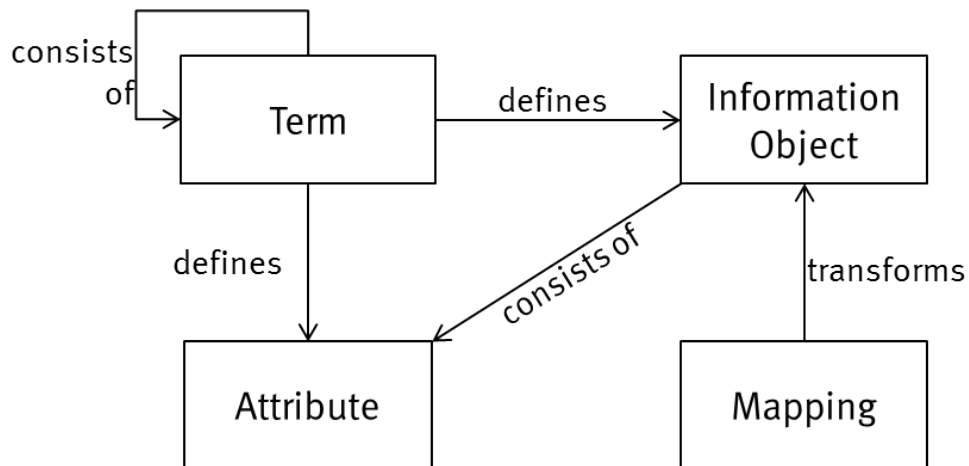


Figure 5: Sub-Model „Information Objects“

GSIM 1.1 contains within the group „Business“ in conjunction with process modeling both „Process Input“ and „Process Output“. The particular input and output objects are specific information objects although a general equivalent object doesn't exist. Furthermore GSIM contains „Rules“ for transformations of objects but it's used primarily for deriving new variables and units (supporting GSBPM Sub-process 5.5).

3. Exchange

The object group „Exchange“ contains information objects, which are closely tied to the exchange of data. Especially the collection of data (e.g. using collection instruments) and the dissemination of products belong to this group.

The category „Instruments“ is especially aligned with the user needs. This category contains the information objects „Questionnaire“, „Question Block“, „Questionnaire Component“, „Question“, „Interviewer Instruction“ and „Statement“.

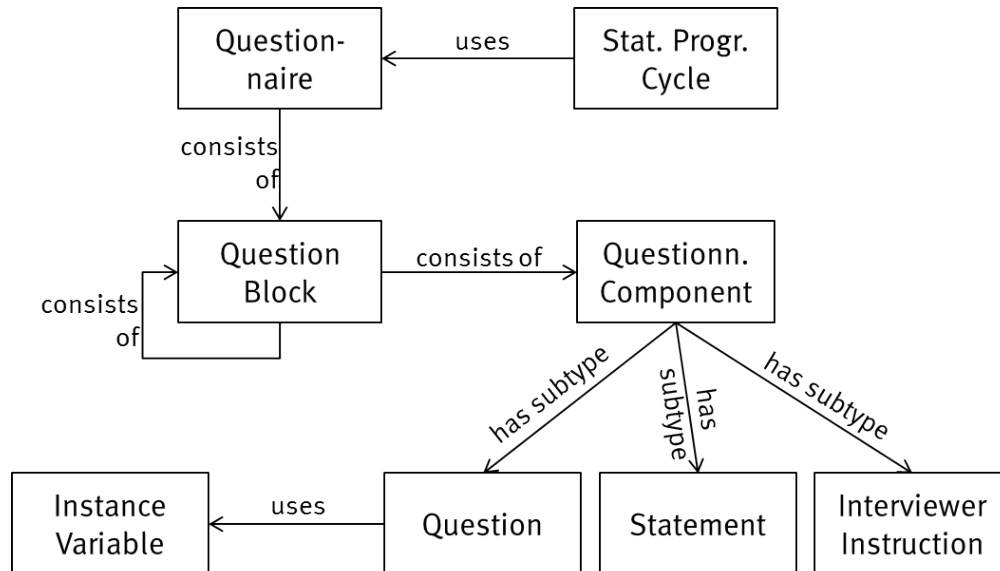


Figure 6: Sub-Model „Instruments“

In this section Germany's core model corresponds to GSIM to a great extent.

4. „Concepts“

The object group „Concepts“ encompasses information objects which contain definitions and descriptions for the comprehension of statistical products for example. The most important information objects are „Unit“, „Variable“ and „Classification“.

The category „Glossary“ comprises definitions and explanations for statistical terms (statistical terms as well as comprehensive terms of certain working groups). The glossary is in relation to the remaining information objects by providing definitions. Potentially the glossary is connected to variables, codes, classification items etc. as well. Essentially the glossary could be modeled completely by just one information object („Term“).

GSIM 1.1 put in place information objects, which are connected to terms, in the group „Concepts“. The information object „Concept“ is connected to a „Concept System“, which differentiates the elements of classifications and code lists. Every element of these gets an unambiguous definition. Generally the connection of the information object „Concepts“ to other information objects could be possible, too.

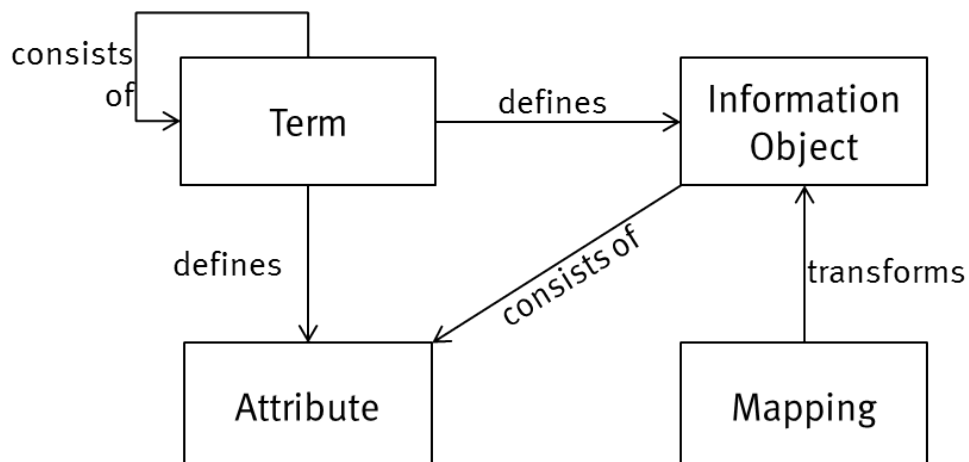


Figure 7: Sub-Model „Glossary“

The category „Variables“ encompasses a sub-model, which is derived mainly from the Neuchâtel Model for variables. Therefore this category contains such information objects like „Variable“, „Represented Variable“, „Conceptual Domain“, „Value Domain“, „Statistical Unit“, „Measurement Unit“, „Quality Value“, „Sign“, „Code“, „Value“ and „Code List“.

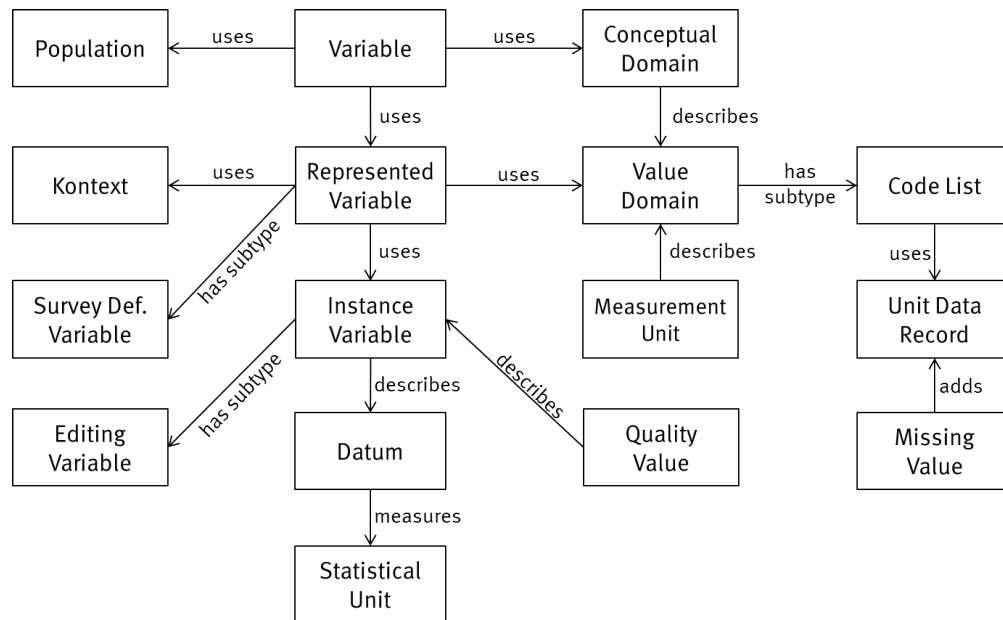


Figure 8: Sub-Model „Variables“

GSIM distinguishes three layers for the modeling of variables within the group “Concepts”:

- „Variable“
- „Represented Variable“
- „Instance Variable“

A „Represented Variable“ is already connected to a „Value Domain“ which is specified by a type and a measurement unit. As with a „Variable“ a description or an enumeration of possible codes is defined. An „Instance Variable“ is connected to a certain „Datum“, which describes a characteristic or observation of a certain „Unit“.

The category „Classifications“ is oriented towards the previous Neuchâtel Model for Classifications and contains especially the information objects „Classification Family“, „Classification“, „Classification Version“ / „Classification Variant“, „Classification Item“, „Classification Level“, „Correspondence Table“, „Map“, „Classification Index“ and „Classification Index Entry“.

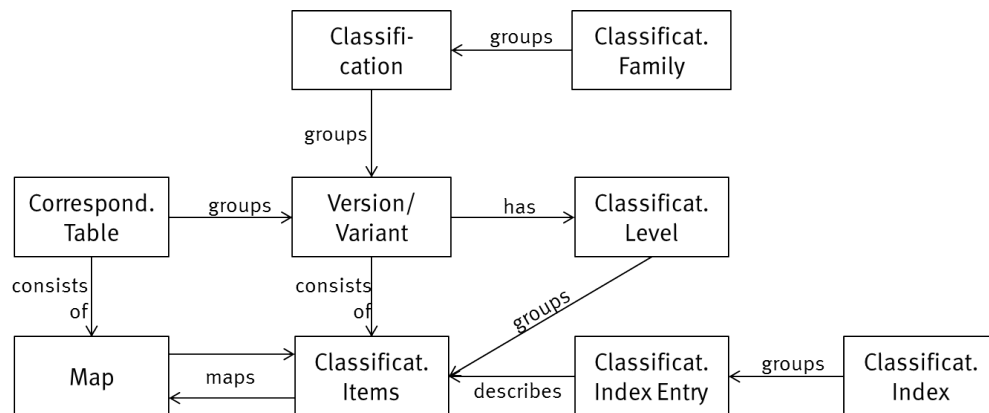


Figure 9: Sub-Model „Classification“

Since GSIM 1.1 is also oriented towards the Neuchâtel Model, the section of classification within the group “Concepts” is mostly compatible to Germany’s core model. “Classification Version” and “Classification Variant” is combined in the information object “Statistical Classification”, though, as Germany’s classification database handle this technically as well.

5. Structure

The object group „Structures“ focuses strongly on data, data flows, data sets and the description of products (content, location etc.) and contains e.g. objects to describe the data structure and data flow but also quality information.

The category “Data Structure” supports external users who are interested in the description of data structures (like researchers who may interested in Scientific-Use- and Public-Use-Files). This category contains especially the information objects „Data Set“, „Data Structure“, „Data Point“ and „Unit Data“.

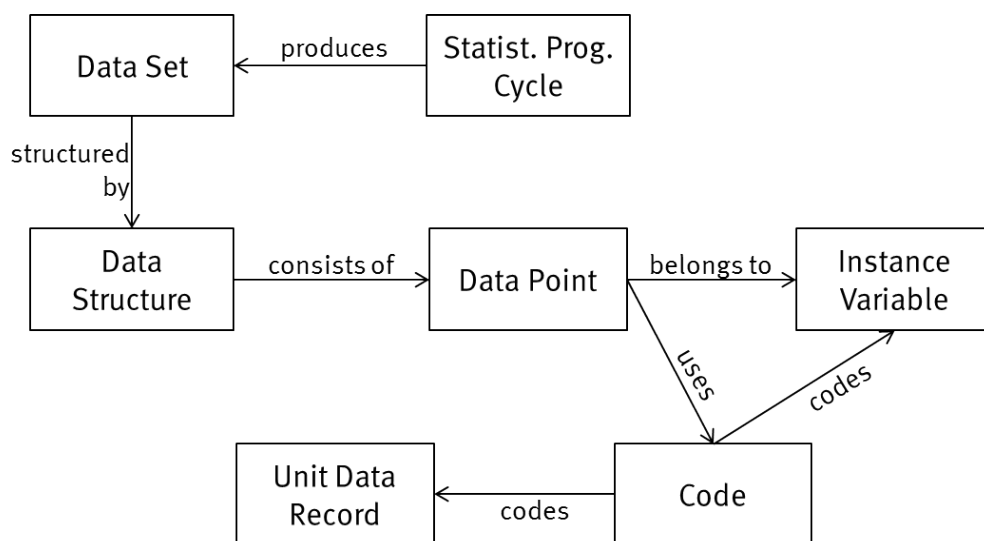


Figure 10: Sub-Model „Data Structure“

GSIM 1.1 describes data structures comprehensively and in detail. Against this Germany’s core model is within this category still in the beginning without much detail. At the moment just a few information objects are sufficient and the compatibility is already in mind though.

Quality information is produced primarily as paradata during the production cycle by the production tools. The core model contains the information objects “Quality Information Template”, “Quality Information”, “Quality Report Template”, “Quality Report” and “Quality Report Provision Agreement” (rather in the group “Exchange”) as seen in Figure 11.

The provider of a quality information can be a person or a production tool.

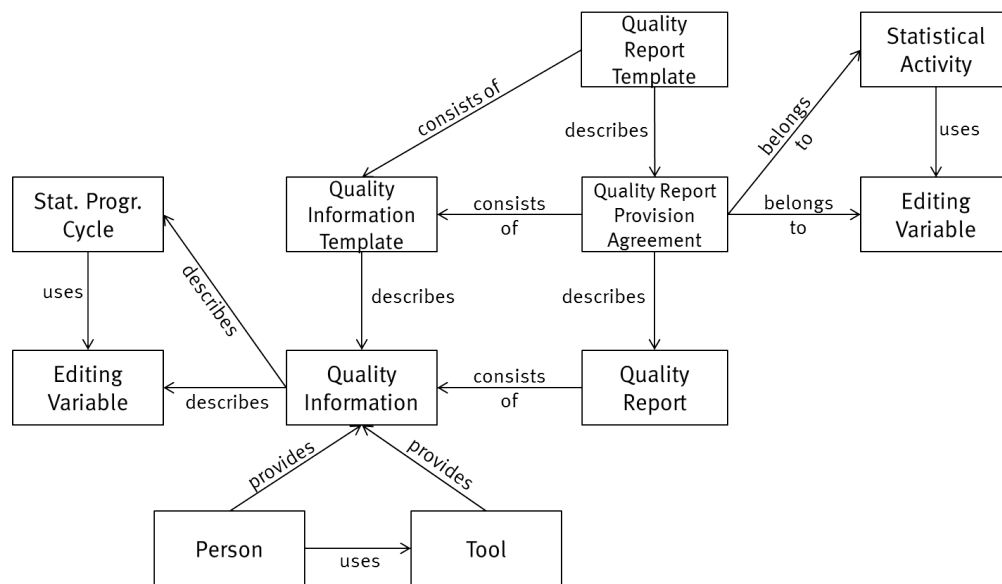


Figure 11: Sub Model „Quality Information“

For modeling quality information GSIM 1.1 uses within the group „Business“ the information object „Process Output“. This object is connected to „Process Step“, which describes technical aspects about the execution of a process step. The „Process Output Specification“ describes the expected outputs for a specific process step.

In addition to the previous version GSIM 1.1 contains several new information objects within the group “Structures” to model the content and structure of referential metadata. Hereby the quality reporting within the ESS (using different report structures like ESMS, ESQRS, SIMS etc.) could be supported effectively. German’s core model seems to be more or less compatible but in the end only the attributes of the objects could create certainty.

Adoption of GSIM

Generally the German Statistical Verbund sees big advantages in adopting and using an internationally standardized information model especially to ensure interoperability with the European Statistical System. Therefore for the development of a comprehensive information model the German Statistical Verbund has decided to take GSIM into account as a reference. Furthermore the information models of the production tools which are already in use have to be considered, too. Fortunately some of the production tools, the variable server and the classification database, have already implemented the Neuchâtel Model, which is part of GSIM, too. Therefore in these sections the compatibility is already noticeable.

System overview

The information model described above in more detail shall be implemented in a metadata management system to support the exchange of information objects among all standardized production tools and systems in the future. The metadata management system will contain interfaces which enables sending and receiving of objects considering appropriate access rights and transformation of data formats.

This way information objects produced in one sub-process (by one tool) can be shared along the production chain and be reused in another sub-process (by another tool). The metadata management system can actually steer the whole production process in this manner dynamically.

The interfaces take care of putting and getting information objects from / to external tools and systems. By making use of adapter (each for every) external systems, which hold and manage certain objects, can be connected to the metadata management system to offer these objects to other tools and systems as well.

At present the metadata management system is still under construction. A small prototype has been developed to prove the concepts, though.

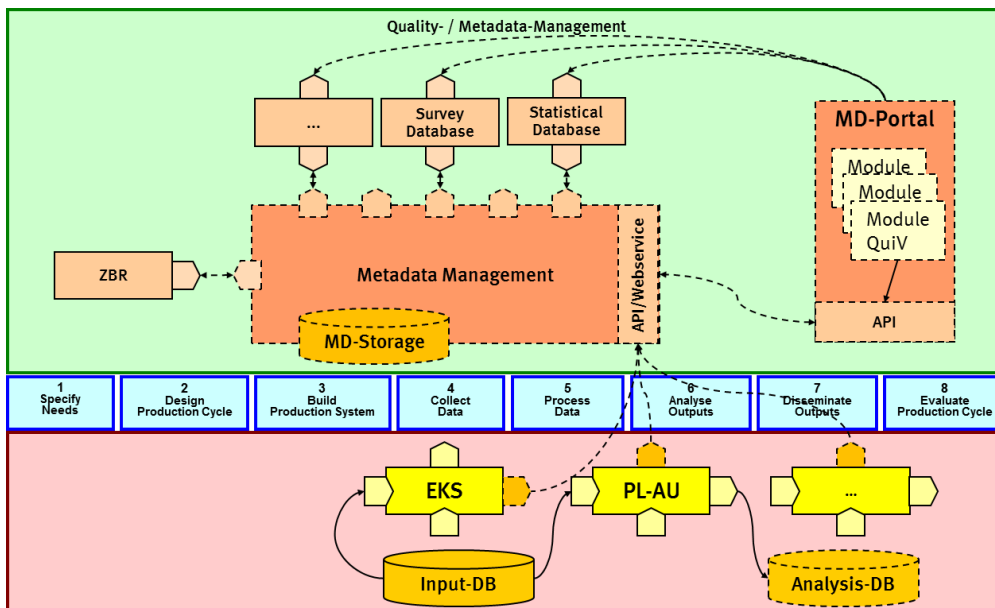


Figure 12: Central role of the metadata management system

Process description that this GSBM implementation supports

In general the use of GSBM with the metadata management system (as a reference for the core model) shall support all sub-process of the GSBPM where metadata is used and/or produced (in the sense of an overarching quality and metadata management). Recent examinations of the existing production infrastructure has shown that the most important tools and systems with their individual information objects especially support the following sub-processes of the GSBPM mainly:

- Phase 2: TP 2.2, TP 2.3, TP 2.5
- Phase 3: TP 3.1, TP 3.2
- Phase 4: TP 4.2, TP 4.3
- Phase 5: TP 5.3, TP 5.5
- Phase 6: TP 6.5
- Phase 7: TP 7.1, TP 7.2

As a result of the examination a process support map has been created (s. Figure 13) which contains for every sub process the supporting production tools and systems and the incoming and outgoing information objects as well.

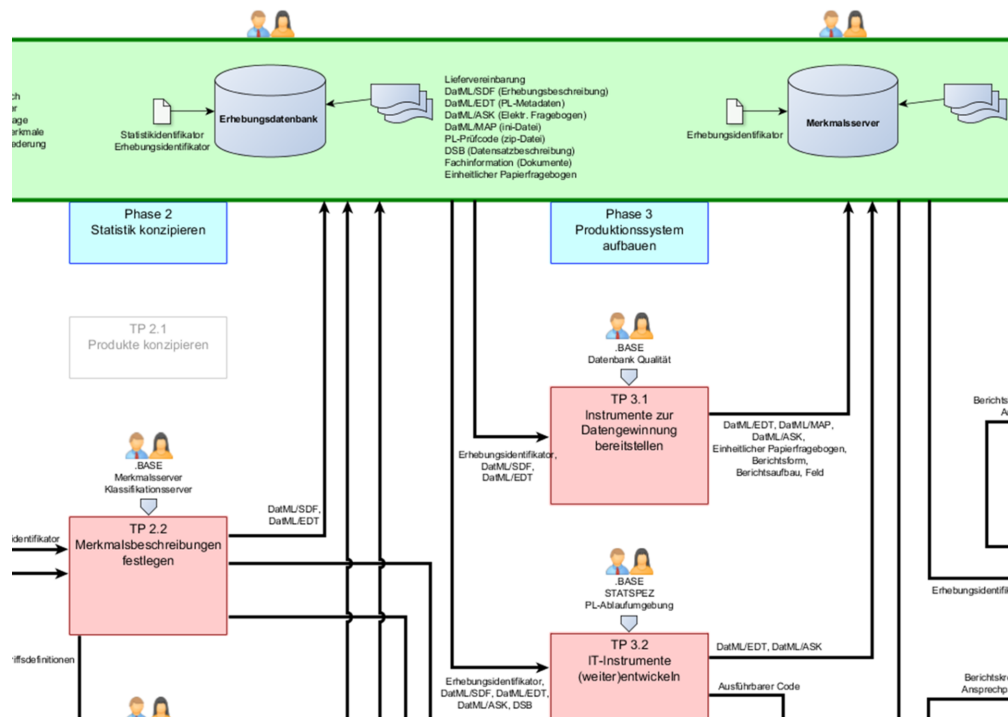


Figure 13: Extract of the process support map

The transformation of the used information objects into and from the standardized information model is one of the first tasks of the metadata management system. The underlying information model is based on GSIM as a reference.

Business case

The future information model of the German Statistical Verbund will be aligned closely to GSIM. The metadata management system which is currently in development will implement the information model. Already existing systems and tools will keep their own information models. The exchange of information objects alongside the production chain will be realized with the support of the metadata management system which transforms the objects into and from the standardized information model. New tools and system may implement the new information model as well.

The development of the metadata management system will be done within the framework of several pilot projects. On one side the statistical portal which shall provide metadata to external users is being drafted for redesign. The metadata management system will care about the provision and transformation of mainly internal metadata to the public under the consideration of access rights and user needs. On the other side the metadata management system will assist the quality management for internal users by providing paradata which are collected by production tools during the statistical production. The paradata will lead to quality information which are joined to quality indicators. The information model will be concentrated initially on these two domains and could be enhanced later on step-by-step.

Relation to other Models

The German Statistical Verbund has adopted the GSBPM and has customized the third level to fit the German needs and reflect the national production environment. The development of the national information model will be aligned very close to GSBPM and GSIM as well.

The German Statistical Verbund has implemented the SDMX standard for the exchange of statistical data especially with international organizations like Eurostat. Therefore the information model will also bear in mind SDMX to utilize international harmonized metadata efficiently and simplify the exchange. Since GSIM also take care of this we don't expect considerable problems with that.

The classification server supports for the import and export of classifications the XML-based format CLASET as Eurostat's RAMON is doing that as well.

However, DDI is not supported by the German Statistical Verbund officially yet.

Design

The information model which is currently still under construction will be implemented by the German Statistical Verbund primarily in the metadata management system. After becoming a standard the information model may be mandatory for new production tools and systems. The already existing tools and systems should be references to guide the design of the new information model to get the largest possible compatibility to information models already in use. The compatibility will be technically accomplished by functionalities which support the mapping and transformation of information objects. Furthermore the guidance by international standards like GSIM is important and future-oriented and may be a perfect base for these transformations.

The information model should be expandable on demand and needs in a step-by-step approach. At first the support of the statistical portal to provide metadata for external users and the support of the quality management for internal users is most important. Hereby the suitability of the information model can be proofed in practical use at an early stage of development. The ongoing production processes should not be impaired by the introduction of the new information model.

Licensing

The German information model is still in construction. Purpose is just the internal use within the German Statistical Verbund. Therefore there are no deliberations about licensing yet.

New Information Objects and/or new specialisations of GSIM Information Objects

In the description of Germany's work to develop an information model build on GSIM mentioned above a comparison to GSIM was already made. For example the modeling of production tools was added in Germany's core model. Production tools can initiate certain activities as people can do. In doing so the correct assignment of access rights to the necessary resources is very important.

Further supplements which should enhance the GSIM in future versions are not noticeable though since the information model of the German Statistical Verbund is still in progress.

Lessons learned

Germany wasn't involved in the development of GSIM but tried to use it at early stages. The use of previous versions of GSIM wasn't that easy though. Sometimes the meaning of certain information objects wasn't sufficiently clear. The proper usage is practically only possible if the related attributes are defined and well-known.

Germany was interested in the usage of GSIM at an early stage of development. With the dissemination of version 1.0 the adoption of the model has been started. The update to version 1.1 has brought along more changes than expected and therefore the adoption process has to be refined to a larger extent than favored. To keep the compatibility the adoption of an internationally standardized model necessitates the consideration of nearly the newest version and the early adoption of updates, too. This could be very time-consuming if updates are too often or too comprehensive.

One of the biggest advantages of the alignment of GSIM was the immediate availability of a comprehensive foundation for a core information model which can be customized to fit the own needs. Furthermore there is already quite a few work done to connect the GSIM to the process model GSBPM.

Suggestions for changes to GSIM

Changes to GSIM should be done rather infrequently and deliberately (similar to the GSBPM) to avoid too much efforts to adapt the changes in the national context. At this time Germany has no concrete suggestions for changes to GSIM.

Microsoft Excel Tabelle GSIM_classification_German
y.xlsx
Use of the Classification Model in Germany

10 Apr, 2015
by Sven
Claussen

Links:

[KlassService](#)

Attachments