

December 2006

Neuchâtel Terminology Model

PART II: Variables and related concepts

object types and their attributes

Version 1.0

Statistics Netherlands

Ad Willeboordse (AWLE@cbs.nl)

Statistics Norway

Anne Gro Hustoft (anne.gro.hustoft@ssb.no)

Jenny Linnerud (jenny.linnerud@ssb.no)

Statistics Sweden

Klas Blomqvist (Klas.Blomqvist@scb.se)

Swiss Federal Statistical Office

Dominique Held (dominique.held@bfs.admin.ch)

Run Software-Werkstatt

Reinhard Karge (reinhard.karge@run-software.com)

Nadja Duhamel (nadja.duhamel@run-software.com)

US Bureau of Labor Statistics

Dan Gillman (Gillman.Daniel@bls.gov)

Content

Summary.....	5
Scope.....	5
1 Introduction.....	6
1.1 Background.....	6
1.2 Context and purpose.....	6
1.3 Summary of main concepts.....	7
1.4 Other terminologies.....	8
1.5 Implementation.....	9
1.6 Layout of the terminology.....	9
1.7 Object types.....	10
2 Metadata object types by content.....	13
2.1 Data item definition.....	13
2.1.1 _Data item.....	13
2.1.2 _Object.....	15
2.1.2.1 _Elementary object.....	16
2.1.2.2 _Aggregate object.....	16
2.1.3 _Variable.....	17
2.1.4 _Time.....	17
2.1.5 _Value.....	18
2.2 Statistical activity and related concepts.....	18
2.2.1 Statistical activity family.....	19
2.2.2 Statistical activity.....	20
2.2.3 Statistical activity instance.....	21
2.3 Variable structure.....	24
2.3.1 Variables and related concepts.....	24
2.3.1.1 Concept family.....	24
2.3.1.2 Conceptual variable.....	25
2.3.1.3 Statistical characteristic.....	26
2.3.1.4 Statistical unit type.....	27
2.3.1.5 Fixed characteristic.....	30
2.3.1.6 Object variable.....	31
2.3.2 Contextual variables.....	33
2.3.2.1 Contextual variable.....	33
2.3.2.2 Register variable.....	35
2.3.2.3 Cube variable.....	36
2.3.3 Value domains and measurement units.....	36
2.3.3.1 Conceptual domain.....	37
2.3.3.2 Value domain.....	39
2.3.3.3 Value domain item.....	41
2.3.3.4 Measurement unit type.....	43

2.3.3.5 Measurement unit.....	44
2.4 Data structure.....	45
2.4.1 Unit types, registers, cubes and tables.....	45
2.4.1.1 Register unit type.....	45
2.4.1.2 Cube unit type.....	46
2.4.1.3 Aggregated unit type.....	46
2.4.1.4 Elementary unit type.....	46
2.4.1.5 Collection unit type.....	47
2.4.1.6 Matrix.....	47
2.4.1.7 Register.....	48
2.4.1.8 Initial observation register.....	49
2.4.1.9 Final observation register.....	50
2.4.1.10 Cube.....	50
2.4.1.11 Table.....	52
2.4.2 Populations, frame and sample.....	53
2.4.2.1 Population.....	54
2.4.2.2 Target population.....	54
2.4.2.3 Frame population.....	55
2.4.2.4 Survey population.....	56
2.4.2.5 Frame.....	56
2.4.2.6 Sample.....	57
2.4.2.7 Population instance.....	58
2.5 Related concepts.....	58
2.5.1 Foreign Concepts.....	58
2.5.1.1 Classification family.....	59
2.5.1.2 Classification.....	59
2.5.1.3 Classification version.....	59
2.5.1.4 Classification level.....	59
2.5.1.5 Classification item.....	59
2.5.1.6 Subject area.....	59
2.5.1.7 Data collection.....	59
2.5.1.8 Data collection element.....	59
2.5.1.9 Table instance.....	60
2.5.1.10 Record type.....	60
2.5.1.11 Dataset.....	60
2.5.1.12 Footnote.....	60
2.5.1.13 Keyword.....	60
2.5.2 Administrative Concepts.....	61
2.5.2.1 Registered item.....	62
2.5.2.2 Person.....	63
2.5.2.3 User group.....	63
2.5.2.4 Organisation.....	63
2.5.2.5 Update information.....	64
2.5.2.6 Publication reference.....	64
3 Metadata object types by concept levels.....	65
3.1 Conceptual level.....	66
3.2 Contextual level.....	68

3.3 Instance level.....	70
4 References.....	72
5 Annexes.....	73
5.1 Annex I - Notation and multiplicity in figures.....	73
5.2 Annex II - Defining terminology models.....	74
5.2.1 Concept.....	75
5.2.2 Object type	75
5.2.3 Characteristic.....	75
5.2.4 Hierarchy.....	76
5.2.5 Category.....	76
5.3 Annex III - Comparison with ISO/IEC 11179 Metamodel.....	78
5.4 Annex IV - Examples of concepts listed alphabetically by concept.....	81

List of figures

Figure 1 Variables and related concepts.....	10
Figure 2 Data item components and related metadata objects.....	13
Figure 3 Part of a statistical table, showing core economic data for several branches of industry in Norway.....	14
Figure 4 Statistical activity and related concepts.....	19
Figure 5 Variables and related concepts.....	24
Figure 6 Contextual variable.....	33
Figure 7 Value domains and measurement units.....	37
Figure 8 Registers and Cubes.....	45
Figure 9 Population.....	53
Figure 10 Registered item.....	61
Figure 11 Conceptual level.....	67
Figure 12 Contextual level.....	69
Figure 13 Instance level.....	71

Neuchâtel Terminology Model

PART II: VARIABLES AND RELATED CONCEPTS

object types and their attributes

Version 1.0

Summary

In 2004, the Neuchâtel Group issued version 2.1 of the Neuchâtel Terminology Model Classification database object types and their attributes. The main purpose of the work was to arrive at a common language and a common perception of the structure of classifications and the links between them. The present document extends the model with variables and related concepts. The discussion includes concepts like object types, statistical unit types, statistical characteristics, value domains, populations etc. The two models together claim to provide a more comprehensive description of the structure of statistical information embodied in data items.

The Terminology Model is both a terminology and a conceptual model. It defines the key concepts that are relevant for the structuring of metadata on variables and provides the conceptual framework for the development of a database organising that metadata.

The Neuchâtel Terminology Model has a two level structure, consisting at the first level of the object types and, on the second level, the attributes associated with each object type. Both object types and their attributes are defined by a textual description. Since the model belongs to the semantic and conceptual sphere of metadata, it does not include metadata that are related solely to the technical aspects of a variables database. The Neuchâtel Terminology Model is generally applicable and not dependent on IT software and platforms. It may be used in any context where structured information on variables and their related concepts is required.

Scope

This document provides a framework for documenting variables and data sets in statistical offices. Other kinds of organisations should find it useful, too. It is intended for use by metadata experts, methodologists and information technology professionals, and might also be useful for subject matter specialists.

The framework is in the form of a conceptual model. There are no other specific implementation details implied by the model. The conceptual model does not contain every possible characteristic for describing variables and data sets, and other characteristics may be added as necessary in implementations for specific needs.

1 Introduction

1.1 Background

After completion of the Neuchâtel Terminology Model Classification database object types and their attributes [ref. 1], the need was felt to develop a corresponding terminology for another core concept in the modeling of statistical information: the variable. The Neuchâtel Group initially took up this subject in 2003. At that time, the group consisted of representatives of the National Statistical Institutions of Denmark, the Netherlands, Norway, Sweden and Switzerland, and the German company run Software-Werkstatt (developers of the BRIDGE software). Statistics Denmark left the group after the work on classifications was completed. The Bureau of Labor Statistics (USA) joined the group, mainly in view of the tight relationship between the statistics-oriented contextual variable and the more general *Data Element* concept, as introduced in ISO/IEC 11179 [ref. 2].

While working out the Terminology Model for Variables (TMV), it became clear that it would make little sense to confine the scope and thus the structure of the model to the conceptual variable only. Indeed, modelling the variable comes down to establishing its relationships with other core concepts, of which the statistical unit type and its statistical characteristics are the most important. There is also a close link with the classifications issue, as the categories of a classification shape the value domain of a variable. It is also important to take into account the context within which variables apply, i.e. the statistical activity.

1.2 Context and purpose

In spite of its name, the Neuchâtel Terminology Model is not just a vocabulary in the sense of a mere collection of terms related to the conceptual variable. It has been developed for several purposes, using a specific methodology, ordering the concepts in a two-level structure of object types and attributes. On the first level, it specifies the object types, while on the second level, it lists the attributes associated with each object type. These attributes refer both to characteristics of the object types and to relations with other object types.

The object types and attributes for variables and related concepts that are listed and defined in this document refer to conceptual metadata only. They are definitional, content-oriented metadata, that order and describe the meaning of the components making up statistical data.

The two models for classifications and variables can, in combination, be conceived as one integrated model. This integrated model aims at providing a logical structure in which all statistics fit, and to provide a conceptual basis for the various servers (statistical unit, variable, classification, value domain and statistical activity servers) of a statistical office. In doing so, it aims at several additional benefits:

- A tighter and more logical connection between the various stages of the statistical process;
 - A smoother communication between those working in different stages of the statistical process;
 - A smoother communication between those working in different subject-matter areas;
- A better understanding between statisticians and information analysts.

On the other hand, the scope of the model is not concerned with recording *all* the terms applying in the field of statistical modelling. It does not deal with methods or best practices in the development and management of metadata, e.g. naming conventions. Since the Neuchâtel Terminology Model belongs to the semantic and conceptual sphere of metadata, it does not include object types and attributes that are solely related to the technical aspects of a server for variables. *Process* metadata and *technical* metadata are outside the scope¹ of this document.

¹ For process and technical metadata we refer to "Reference Model TM" [ref. 4] as provided by the METANET project of Eurostat.

The next section provides an overview of the model description, which is described in detail in **section 2**, by summarising the main concepts of the model.

1.3 Summary of main concepts

The actual model description starts in **section 2.1** with the **data item definition**. An example² is the statement that *on the 31st of December 2002, in Norway, the number of employees, in establishments in the Norwegian wholesale and retail trade with 30 or more employees, was 77334*. The data item is a composition and instantiation of four of the core concepts of the model, i.e. object, variable, time and value. In section 2.1 these core concepts are briefly explained; in later sections they are further elaborated.

Section 2.2 deals with the notion of a **statistical activity**, the chain of actions taken by a National Statistical Institution in order to produce a collection of data items. This collection generally takes the form of one or more statistical tables, each cell of which holds a data item. The statistical activity *structural statistics on wholesale and retail trade* undertaken by Statistics Norway generates a set of tables containing time series on the *turnover, employment, compensation of employees, investments etc in all branches of the Norwegian wholesale and retail trade sector*.

Various statistical activities belonging to the same subject matter area can be grouped into a statistical activity family. The *structural statistics family* is composed of *the wholesale and retail trade statistics*, together with *statistics for the manufacturing industry, construction industry etc*.

The *annual versions of the wholesale and trade statistics* are statistical activity instances.

Section 2.3 introduces the **variable structure** including the conceptual variable. Applying this concept to the statistical characteristics of a statistical unit type generates the object variable. Applying the conceptual variable of *employment* to the statistical characteristic *having employees* of the statistical unit type *establishment* generates the object variable *number of employees of establishments*. Not all statistical characteristics of a statistical unit type generate object variables; some characteristics are the same for all instances and are therefore called fixed characteristics. For the statistical unit type *hotel*, *kind of activity* is not an object variable, but a fixed characteristic as its values cannot vary for various instances of the statistical unit type *hotel*. In other words, in this example only one category or value domain item of the conceptual domain for the activity classification NACE (classification of economical activities) applies. An object variable and a conceptual domain used in a specific context, i.e. for a specific statistical activity, are called a contextual variable and a value domain.

An object variable may appear in cubes as a cube variable with basically two types:

1. As a classifying cube variable, it subdivides the population of statistical unit types into subpopulations, e.g. *If, in a cube, the total population of establishments is subdivided into subpopulations according to size-classes, the numerical register variable number of employees acts as a classifying cube variable.*
2. As a quantifying cube variable, its values show - after aggregation - in the cells of the cube e.g. *If, in a cube, the total number of employees for a population of establishments is recorded, the numerical register variable number of employees acts as a quantifying cube variable.*

Numerical register variables can be used in both roles in cubes, whereas categorical register variables, like kind of activity, appear in the role of classifying cube variables only in cubes.

In **section 2.4** the model moves from the variable structure to the **data structure**. This describes the way that data are ordered in data collections. Essentially, a data structure can be described from two angles:

1. As the intension³ of a data collection, i.e. the intension of registers, cubes and tables. These can be intensionally defined in terms of statistical unit types and sets of object variables. Registers

² All examples are based on the statistical activity *wholesale and retail trade statistics* of Statistics Norway. However, for reasons of simplification examples may deviate from real practise.

relate to register unit types (elementary units), while cubes relate to cube unit types (aggregated or elementary units).

Example: For the statistical activity wholesale and retail trade statistics of Statistics Norway:

- *The final observation register is based on the elementary register unit establishment and the object variables number of employees and turnover;*
- *The cubes and tables are based on the aggregated cube unit trade sector.*

2. As the extension⁴ of a data collection, i.e. as a set of objects, called the population. As such, a population denotes the extension of a statistical unit type, and is always associated with a particular statistical activity. In different phases of a statistical activity, different types of population apply:

- In the output-design phase, the target population is defined;
- In the data collection-design phase, the frame population and the survey population is defined, taking into account the frame available.

Example: For the statistical activity wholesale and retail trade statistics of Statistics Norway:

- *The target population for the cubes and tables resulting from the wholesale and retail trade statistics in year t is defined as: all establishments existing at any moment in year t , and engaged in wholesale or retail trade as their main kind of activity.*
- *The frame population for the collection of data for the year t consists of all establishments registered as active in the Norwegian Central Register of Enterprises and Establishments in the year t , and classified as NACE-industry divisions 50, 51 or 52, including relevant information about each establishment, such as address etc.*
- *The survey population for the collection of data for the year t consists of all establishments registered as active in the Norwegian Central Register of Enterprises and Establishments in the year t , and classified as NACE-industry divisions 50, 51 or 52*
- *This Central Register is the for the statistical activity mentioned.*

1.4 Other terminologies

There exist a number of terminologies and glossaries dealing with terms and concepts associated with variables. ISO/IEC 11179 is one of the most closely related metamodels.

The ISO/IEC 11179 metamodel (11179-MM) and the Terminology Model for Variables (TMV) touch on similar statistical metadata areas, but from different perspectives. 11179-MM is a general description of data, independent of the subject area. It also supports registration, a methodology for administering content including its quality. The TMV, on the other hand, describes and classifies statistical data. Because its context is statistics, the TMV contains concepts and terminology familiar to statisticians.

Interestingly, the structure of both models is quite similar. This similarity allows the exchange of metadata between ISO/IEC 11179 registries and TMV repositories⁵. Thus, one may generate an ISO/IEC 11179 metadata registry from a TMV compliant metadata system, and vice-versa.

³ Intension - sum of characteristics that constitute a concept.

⁴ Extension - set of objects to which a concept refers.

⁵ Metadata registries and metadata repositories are both databases of metadata. However, a metadata registry also supports the functionality of registration.

Combining the administrative components of ISO/IEC 11179 with the detailed, statistics specific, content definition of TMV provides the basis for a standards-conformant⁶ statistical metadata registry. This covers both registration aspects and rich content definitions for statistical offices.

In view of the close relationship between TMV and 11179-MM, the vocabulary of the TMV has been adapted as much as possible from that of 11179-MM. As the vocabulary of the latter is in turn based on that of ISO 1087-1 [ref. 3], it can be said that the vocabulary of the TMV complies with ISO-norms.

Annex 3 of this document provides a systematic comparison between TMV and 11179-MM.

1.5 Implementation

Both the terminology and the conceptual model are generally applicable and not dependent on IT software and platforms. The conceptual model can be used in any context where structured information on statistical variables is needed.

1.6 Layout of the terminology

Section 2 starts with an overview of the object types listed by content. The following pages contain the list of all object types and their characteristics. The object types are listed in the same order as in the overview. Each object type is defined by a textual description, followed by a list of the characteristics associated with the object type. Each characteristic is also described. A few examples have been added to facilitate understanding. There has been an attempt also to order the characteristics according to some sort of logic and to list them in a consistent way across the object types. If the symbol [--> Object type] appears in a description then this refers to an object type listed and described elsewhere in the terminology.

While object type terms are unique, the name of a characteristic may differ in meaning when it is associated with different object types. Naming conventions are considered to be an implementation detail and are not included in this document. Some of the central object types of the terminology, *e.g. statistical activity and contextual variable*, have quite a number of attributes attached to them. For certain applications some of the attributes will be superfluous. They need not all be used.

Time has not permitted a thorough review of the descriptions. We are aware that they are not consistently of one kind but waver between subject matter oriented and IT oriented language, sometimes they are genuine definitions, sometimes they indicate how the information will appear in the technical application. In spite of good intentions, it has been difficult to keep the conceptual and the implementation levels separate.

Section 3 provides an overview of the object types listed by level of abstraction; conceptual, contextual and instance levels.

The first annex provides an explanation of the notation and multiplicity used in all our figures. The second annex provides detailed information on terminology models. The third annex contains a comparison of part of the Neuchâtel Terminology Model for variables and related concepts with the metamodel specified in Part 3 (2nd edition - 2003) of ISO/IEC 11179. The fourth annex provides a list of examples for most of the object types. The list is sorted alphabetically by object type term.

⁶ Conformance is a specific term in standards work. It means, roughly, any implementation of a standard is conformant if it satisfies all the requirements in the standard.

1.7 Object types

The following figure is a copy of the figure in section 2.3.1 and gives an overview of the main object types and relationships for variables and related concepts. These object types are more extensively described in section 3. See the first annex for an explanation of the notation and multiplicity used in all our figures.

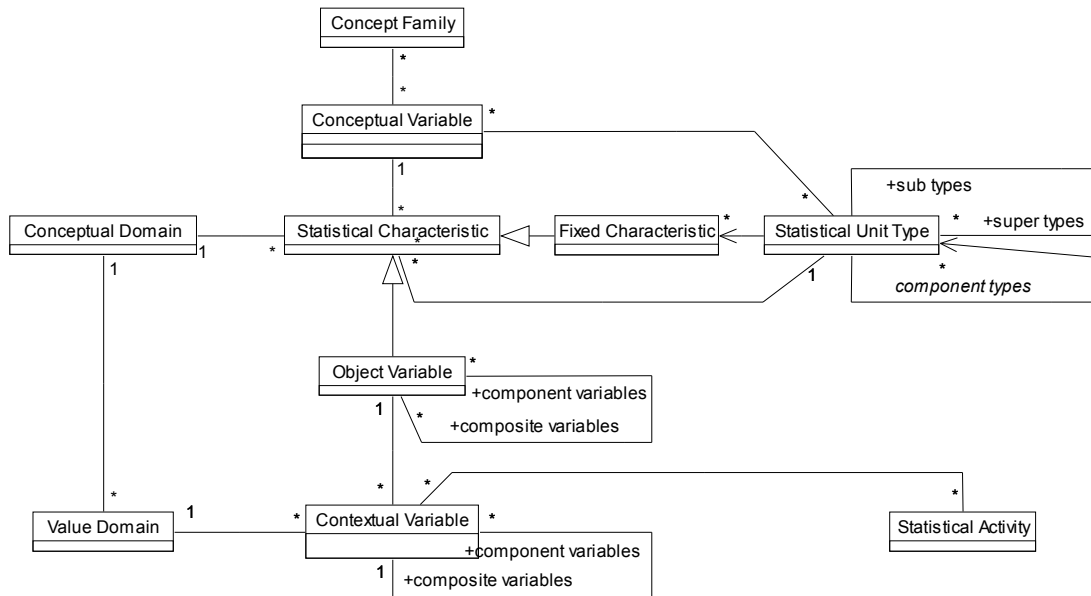


Figure 1 Variables and related concepts

<p>Statistical activity: Within a statistical office, a statistical activity is an activity that produces collections of statistical data.</p>
<p>Concept family: A concept family comprises a number of conceptual variables, which are related from a certain point of view.</p>
<p>Conceptual variable: The conceptual variable defines the general concept of a variable, independent of its relation to a statistical unit type or its use in a statistical activity.</p>
<p>Statistical characteristic: Statistical characteristics are defined as characteristics of a statistical unit type. A statistical characteristic can either be fixed (fixed characteristic) or variable (object variable).</p>
<p>Statistical unit type: A statistical unit type describes a class of statistical objects and its relation to super class(es) - generalisations - and sub-classes - specialisations.</p>
<p>Fixed characteristic: A fixed characteristic is a statistical characteristic that describes an essential characteristic of a statistical unit type and is part of its definition.</p>
<p>Object variable: An object variable defines the concept of a variable in connection with a defined statistical unit type and a conceptual domain.</p>
<p>Contextual variable: A contextual variable defines a variable in the context of a statistical activity.</p>
<p>Conceptual domain: Conceptual domains can be enumerated (listed) or non-enumerated (not listed) i.e. the conceptual domain defines a set of categories or includes a declaration of the type of measurement unit that can be used for certain statistical characteristics (object variables).</p>
<p>Value domain: A value domain defines the specific valid values (domain) for a contextual variable.</p>

2 Metadata object types by content

This document describes metadata object types from two viewpoints, reflecting two different ways in which to structure metadata. The first and main description is by content. In the second view, added at the end of the document (section 3), metadata object types are presented on three different levels: conceptual, contextual and instance levels.

This section describes metadata object types ordered by content. The section refers to five categories of metadata object types (data item definition, statistical activity, variable structure, data structure and finally, related concepts). The most important categories are variable structure and data structure. The first category (data item definition) introduces a definition of data, which metadata refers to. The last category (related concepts) refers to metadata object types not described in detail in this document.

2.1 Data item definition

This section introduces the notion of a data item, which is the most elementary piece of statistical information presented in statistical tables. The data item is data rather than metadata. The definition of data items is an essential part of variable definitions. Breaking down the data item into its constituents illustrates how metadata objects relate to data item components. Since metadata is related to data directly or indirectly, this section also shows the essential links between data and metadata, although the metadata object types have not yet been defined.

To be able to differ between metadata object types and data item components, all data related concepts start with a '_' (e.g. `_Variable`).

The following diagram provides an overview of data item components and related metadata objects, which are described later on. All data item components can be described by means of metadata.

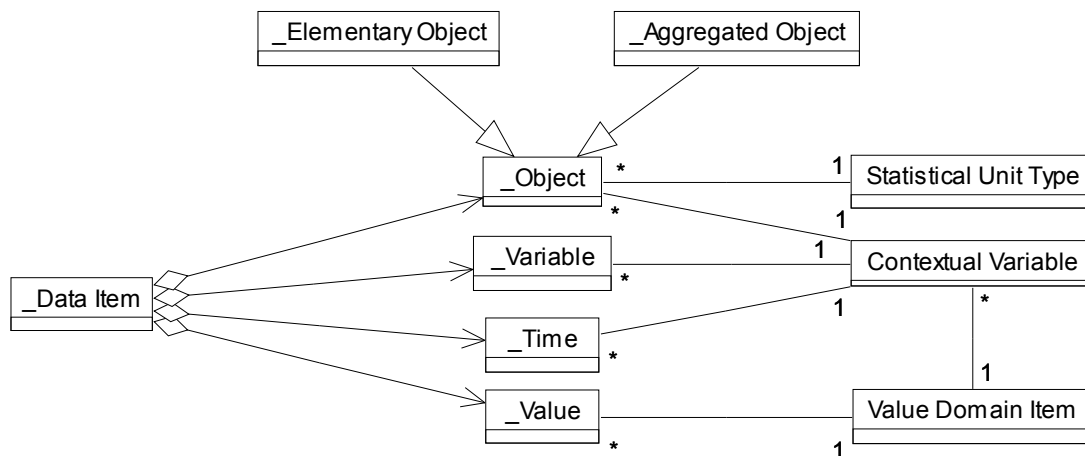


Figure 2 Data item components and related metadata objects

2.1.1 `_Data item`

A data item is an instantiation (occurrence) of a context variable. A data item is considered as an atomic data value (number or text) of a variable that is related to an elementary or aggregated object at a certain point of time or for a given time interval. Thus, a data item is described by a particular combination of the dimensions object, variable, time and value, where the latter is derived from the value domain applying for the context variable.

The components of a data item together define the content of a cell in a statistical table, being the final product of a statistical activity.

<i>Branch of industry</i>	<i>Year</i>	<i>Number of enterprises</i>	<i>Number of employees</i>	<i>Turnover (mill.NOK)</i>	<i>Gross investments (mill.NOK)</i>
<i>Manufacturing industry</i>	2002	10 205	278 773	499 166	17 908
	2003	9 718	264 299	497 903	14 019
<i>Transport and communication industry</i>	2002	22 959	159 477	310 362	13 438
	2003	22 776	154 657	312 663	18 879
<i>Wholesale and retail trade, repair activities</i>	2002	57 395	319 937	877 707	9 462
	2003	57 282	322 230 [§]	891 444	10 125

Figure 3 Part of a statistical table, showing core economic data for several branches of industry in Norway

The data item in the marked cell (§) is defined by:

- *At the elementary object type level, the statistical unit (sub) type wholesale and retail trade, repair activities enterprise⁷*
- *At the aggregate object type level, the statistical unit (sub) type wholesale and retail trade, repair activities branch*
- *The category “322 230” of the context value domain of the (quantifying) variable number of employees*
- *The category 2003 of the value domain for the variable “time”*

Any subset of data item components can be defined as metadata on a conceptual level defining a single or a group of potential observations. Usually, metadata conceptually describes the object(s), variable and/or time component for the objects to be observed. It is also possible to define only object and variable on the conceptual level and observation time and value on data level. However, situations may also arise in which the object, variable and value are defined conceptually and the time is observed, *e.g. observe when a person (object) reaches the size (variable) of 1.50 m (value)*.

Thus, it is a matter of perspective, what is considered as data and as metadata. Nevertheless, a data item is considered as complete, when it is designated by its four components.

There is no doubt that the value is one component of a data item, but there are many discussions, whether the three other item identifying components are required or not. In some cases, the time component seems to be redundant especially when the value does not vary significantly over time, (*e.g., the mass of the earth*). Other examples may show that the outcome of an observation depends on the observer, i.e. three different observers may produce three different values, in which case the data item requires another identifying component. For statistics as well as for many other areas, time, variable (concept) and object instance seem to be necessary and sufficient.

⁷ One can also state that *Enterprise* is the statistical unit type here. In that case, *branch of industry* has to be added as a (classifying) variable, with *wholesale and retail trade, repair activities* as the category of its value domain/

Characteristic	Description
_Object	The object is the thing or event for which the value for a particular variable is observed. [--> _Object]
_Variable	The variable describes the concept of the observation that the data item results from. [--> _Variable]
_Time	The time refers to a time point or time period on or during which the data item was observed. [--> _Time]
_Value	The value is the observed value of the contextual variable for the data item at the referred time. [--> _Value]

2.1.2 _Object

The object may refer to a single thing, in which case it is an elementary object (*e.g. the establishment Ericsson AS Grimstad (Norway)*) or it may refer to a set of things, in which case it is an aggregated object (*e.g. the trade sector, defined as the set of establishments with trading as major activity*). Each object in a data item is associated with one statistical unit type.

In contrast to an object type (*trade establishment*), the object instance (*Ericsson AS Grimstad (Norway)*) is subject to observation (or estimation), i.e. the object is specific, while the object type is abstract and actually is a term denoting the collection of all specific objects complying with its definition. Any type of designation may identify the object that a data item is related to.

Characteristic	Description
Identifier	An object can have one or more specific characteristics, which formally identify the object instance. Typically, those are identifying numbers or other administrative characteristics, which have been introduced explicitly to identify an object instance.
Statistical unit type	An object is associated with a statistical unit type, which reflects the way an object is viewed. <ul style="list-style-type: none"> - <i>One may see the object Angela Merkel as a citizen or as a chancellor</i> - <i>One may see the object Ericsson AS Grimstad as an employer, a tax payer or as a trader</i> [--> Statistical unit type]
Components	A component is a part of an object that may be an object in its own right <i>The establishment Ericsson AS Grimstad is a component object of the enterprise Ericsson AS Norway</i> [--> _Object]

Characteristic	Description
Composite	<p>A composite object is a combination of object instances (components), which forms an object in its own right</p> <p><i>The enterprise Ericsson AS Norway is a composite object of the establishment Ericsson AS Grimstad and other establishments.</i></p> <p>[--> _Object]</p>
Contextual variable	<p>An object is associated with a contextual variable.</p> <p><i>In our general example the object Ericsson AS Grimstad is associated with the variable number of employees and with the variable kind of activity:</i></p> <p>[--> Contextual variable]</p>
Characteristics	<p>An object (instance) has one or more characteristics (properties on instance level). Characteristics of an object are used to describe or distinguish the object.</p> <p><i>The object Ericsson AS Grimstad, belongs to a Swedish multinational, sells machinery, employs 83 staff and is located in Grimstad</i></p>

2.1.2.1 _Elementary object

An elementary object is a particular identifiable thing *e.g. the establishment Ericsson AS Grimstad (Norway)* or event. Each elementary object can belong to one or more object types (statistical unit types).

Elementary objects are subject to observation (or estimation). Object instances which result from observations of objects (and later estimations) rather than from aggregating details of observed object, are considered as elementary object instances. Elementary objects are, in many cases, objects like *persons or enterprises*, which one can grasp and give names to, but also more abstract objects like *accidents or holidays or jobs*.

Characteristic	Description
_Object	<p>An elementary object is a special type of object and inherits all the characteristics of object.</p> <ul style="list-style-type: none"> - <i>at type level: establishment</i> - <i>at instance level: the establishment Ericsson AS Grimstad</i> - <i>at instance level: the accident with the Estonia ferry</i> <p>[--> _Object]</p>

2.1.2.2 _Aggregate object

Aggregate objects are object instances that define a collection of elementary or aggregate objects. Values for aggregate contextual variables are usually not observed directly but obtained from aggregation.

Thus, Berlin can be considered as an aggregate object of all persons living in the region Berlin. Likewise, the Norwegian trading sector is the aggregate object of all establishments in Norway with major activity trading

Typically, categories defined in a classification define aggregate objects when applied to a certain population, *e.g. trading*. Whether an object is an elementary or an aggregate object depends on the

view or context. Thus, an establishment can be considered as elementary object in a register but also as an aggregate object in a cube for the people working for the enterprise.

Aggregate objects are usually associated with aggregated data (e.g. *the total turnover and number of employees of the Norwegian trade sector in 2002*).

Characteristic	Description
_Object	An aggregate object is a special type of object and inherits all the characteristics of object. [--> _Object]

2.1.3 _Variable

A variable defines the concept of an observation (or measurement) for a given statistical unit type. The variable describes the concept of the observation that the data item results from. Thus, the variable is always associated with a contextual variable that describes the concept of the variable in the context of a particular statistical activity.

Characteristic	Description
Contextual variable	The variable describes the concept behind the observed value, which is described as contextual variable (or as object variable or as conceptual variable on more abstract levels). <i>Number of employees in the Norwegian trade sector as observed in the statistical activity structural statistics on wholesale and retail trade by Statistics Norway is a contextual variable derived from the conceptual variable employment</i> [--> Contextual variable]

2.1.4 _Time

Each data item or observation has a time reference, i.e. each data item is associated with a time value (time point or period). This is not always obvious, since the time value might be hidden. In some cases, more than one time point is required, especially, when recording values (value of import or export, income etc.). Since values are measured in currency units, which change in itself, the measured value will change depending on the value of the currency at a given time point.

When considering the event time, which is the time at which the value is supposed to be a characteristic of the object instance, and the observation time, which is the time at which the observation was made, event time and observation time need not to be identical. When the observation time affects the value, i.e. when the value measured depends on the time that the observation was made, the observation time becomes an additional identifying component for the data item.

Sometimes, all data items in an observation register refer to the same observation time, in which case the time is usually stored as an attribute of the data collection and consequently of all contextual variables belonging to this collection. In other cases, data items for a contextual variable refer to the same time, in which case the time is an attribute of the data collection element. In other cases, all data items for an object are associated with a time value or attribute (time series).

In any case, time is present and is important information for identifying and interpreting a data item.

Examples:

- *Number of employees in the Norwegian trade sector on December 31, 2002*
- *Turnover of the Norwegian trade sector in the year 2002.*

Characteristic	Description
Contextual variable	<p>An observation is made at a certain time point for a certain time point or period. Each data item must be associated directly or indirectly with an observation or event time. When the time point or period is defined as an attribute in the register (time column), it is associated with a contextual variable describing the time value.</p> <p>Observation time is time (date) you observe/measure a phenomenon.</p> <p>Event time is time (date) the phenomenon actually happened.</p> <p>Registration time is time (date) you record the event in a register.</p> <p>In chronological order, these would be event time, observation time and registration time. In some cases, it may be practical to use year rather than date or time.</p> <p>[--> Contextual variable]</p>

2.1.5 _Value

The value describes the quality or quantity observed for a contextual variable. The value is the observed value of the contextual variable for the data item at the referred time.

Characteristic	Description
Value domain	<p>A value belongs to a value domain.</p> <p><i>The value observed for the number of employees on the 31st of December 2002, in establishments in the Norwegian wholesale and retail trade with 30 or more employees, was 77334. This value refers to one item in the value domain applying, i.e. all non-negative integers.</i></p> <p>[--> Value domain]</p>

2.2 Statistical activity and related concepts

Within a statistical office, a statistical activity is an activity that produces collections of statistical data. A statistical activity might occur one or several times, at regular or irregular intervals, and is therefore an overall concept for all of its occurrences, which are called statistical activity instances. Several statistical activities might be grouped together in a statistical activity family.

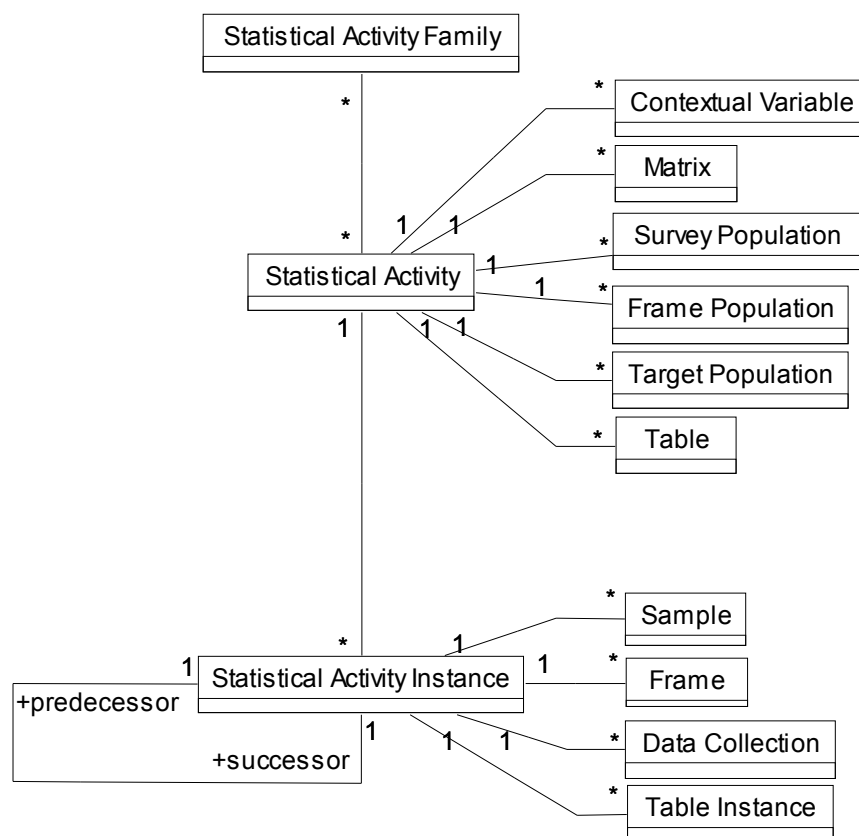


Figure 4 Statistical activity and related concepts

2.2.1 Statistical activity family

A statistical activity family groups statistical activities together according to a systematic definition within a statistical office. A statistical activity family could be a subject matter area.

Example:

The family of structural statistics in Statistics Norway is composed of the wholesale and retail trade statistics, together with annual statistics for the manufacturing industry, construction industry etc.

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the statistical activity family. This may typically be an abbreviation of its title or a systematic number.
Title	The title is a short multilingual label for the statistical activity family.
Description	General multilingual description of the statistical activity family, including its purpose and the type of activities collected in the statistical activity family.
Activities	A statistical activity family refers to a number of statistical activities. [--> Statistical activity]

Characteristic	Description
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.2.2 Statistical activity

Within a statistical office, a statistical activity is an activity that produces collections of statistical data. A statistical activity is often related to a subject area, and can be grouped together with other statistical activities into a statistical activity family.

Under collections of statistical data (data collections) are meant the different kinds of data that can be produced at the different stages of the statistical production process: initial observation registers, final observation registers, cubes, tables or publications. The intension of the data collected or analysed in a statistical activity is described in terms of registers and cubes. The extension of the data collected or analysed in a statistical activity is described by means of population(s).

A statistical activity might comprise various production processes related to the input, throughput or output of statistical data, such as the collection of data (survey), the data cleansing, the combination of data from various sources, data calculations and estimations, analyses, production of tables etc. A statistical activity defines a scope for documenting concepts and production rules for producing statistical data.

A statistical activity might occur one or several times, at regular or irregular frequency, and is therefore an overall concept for these occurrences, which are called statistical activity instances. The statistical activity, while providing conceptually a general frame for all its statistical activity instances, summarises also the results of the related instances, in terms of data collections, data sets or table instances.

Example:

- *The structural statistics on wholesale and retail trade is a statistical activity undertaken by Statistics Norway.*

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the statistical activity. This may typically be an abbreviation of its title or a systematic number.
Title	The title is a short multilingual label for the statistical activity.
Description	Short multilingual description of the statistical activity. This summarises the detailed description.
Detailed description	Detailed multilingual description of the statistical activity. This describes the actions performed within the statistical activity.
Legal base	Indicates that the statistical activity is covered by a legal act or by some other formal agreement.
Activity start	The date that the statistical activity started. The date can be expressed as text, since it might not be an exact date.
Activity end	The date that the statistical activity stopped. The date can be expressed as text, since it might not be an exact date.

Characteristic	Description
Methodological descriptions	Multilingual description of the methods used for the statistical activity.
Current instance	One of the instances of the statistical activity can be assigned as the currently valid instance. This is usually the latest instance that has been produced. [--> Statistical activity instance]
Currently active	Indicates whether the statistical activity is currently active (produces data) or inactive (stand by, stopped).
Family	A list of statistical activity families associated with the statistical activity. [--> Statistical activity family]
Contextual variables	A number of contextual variables associated with statistical activity. A contextual variable belongs to one and only one statistical activity. [--> Contextual variable]
Matrices	This is a collection of matrices (registers and cubes) that are used within the statistical activity. Thus, it indirectly defines the intensional dimension of the statistical activity. [--> Matrix]
Target populations	A number of target populations may be defined for the statistical activity. [--> Target population]
Frame populations	A list of frame populations associated with the statistical activity. [--> Frame population]
Survey populations	A number of survey populations may be defined for the statistical activity. [--> Survey population]
Frames	A list of frames associated with the statistical activity. [--> Frame]
Tables	A number of tables associated with statistical activity. [--> Table]
Activity instances	A statistical activity has one or more statistical activity instances that describe the processes and results for a certain time point or period of the statistical activity. [--> Statistical activity instance]
Footnotes	A number of footnotes or remarks may be defined for the statistical activity. [--> Footnote]
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.2.3 Statistical activity instance

A statistical activity instance describes the instantiation of a statistical activity at a certain time point or period. A statistical activity may refer to one or more activity instances. Within an activity instance,

instantiations of the concepts defined in the statistical activity are created. Thus, the activity instance instantiates populations as population instances, tables as table instances or registers and cubes as data collections. Abstract rules or methods are described as particular processes and their implementations.

While the statistical activity describes, what kind of data has to be produced for which population, and how it has to be produced, the activity instance describes one or several particular production process(es) and its results. Thus, the statistical activity instance may register irregularities or particularities related to the different processes of production of statistical data.

Main resources of a statistical activity instance are data collections and table instances. Data collections are described by matrices (registers and cubes), which define the intension of the data collection, and by population instances, which define the extension of the data collection.

Example:

- *The 2002 survey is an instance of the statistical activity annual structural statistics on wholesale and retail trade.*

Synonyms: Survey instance, Activity instance

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the statistical activity instance. This may typically be an abbreviation of its title or a systematic number.
Title	The title is a short multilingual label for the statistical activity instance.
Description	Multilingual description of the statistical activity instance. This describes the actions performed within the statistical activity instance and the results.
Legal base	Indicates that the statistical activity instance is covered by a legal act or by some other formal agreement.
Methodological description	Multilingual, detailed description of the methods used for the statistical activity instance.
Milestones	Statistical activity instances can be divided in several sections that are marked by milestones. Milestones describe scheduled tasks and dates.
Current instance	Indicates whether the statistical activity instance is the currently valid instance.
Activity	Each statistical activity instance belongs to one and only one statistical activity. [--> Statistical activity]
Successor	When the statistical activity instance has a successor instance, this instance is referenced here. [--> Statistical activity instance]
Predecessor	When the statistical activity instance is a successor of another instance, this instance is referenced here as predecessor. [--> Statistical activity instance]
Populations	One or more population instances can be referenced for defining the instantiation of the frame population (<i>e.g. for persons, households and buildings</i>). [--> Population instance]

Characteristic	Description
Frames	<p>Within a statistical activity instance one or more frames can be defined, (<i>e.g. person or business register, list of phone numbers, etc.</i>).</p> <p>[--> Frame]</p>
Sample	<p>Within a statistical activity instance one or more samples can be defined.</p> <p>[--> Sample]</p>
Data collections	<p>A number of data collections may refer to the statistical activity instance.</p> <p>[--> Data collection]</p>
Table instances	<p>A number of table instances may refer to the statistical activity instance.</p> <p>[--> Table instance]</p>
Registration	<p>The registration provides administrative information in accordance with ISO/IEC 11179.</p> <p>[--> Registered item]</p>

2.3 Variable structure

Definition of metadata related to variables requires a number of metadata objects describing variables of different kind and on different conceptual levels. This section describes variables on different conceptual levels regardless on their usage within specific data structures.

2.3.1 Variables and related concepts

Variables can be defined on two levels of abstraction:

1. Variables defined as concepts outside any specific context (conceptual variable, object variable)
2. Variables defined in the context of specific statistical activities (contextual variables)

Context independent conceptual variables provide structure to variable definitions and support standardisation and harmonisation of variables. Context related variables are describing specific conceptual variables as defined in a statistical survey or analysis.

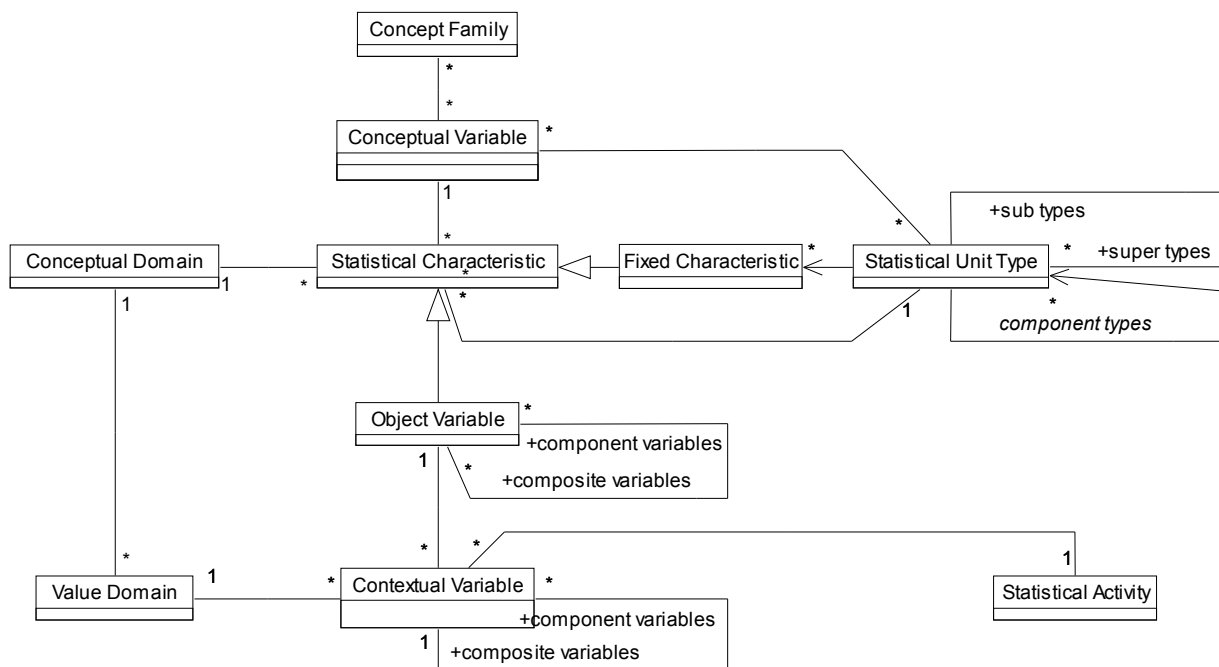


Figure 5 Variables and related concepts

2.3.1.1 Concept family

A concept family comprises a number of conceptual variables, which are related from a certain point of view.

Example:

The conceptual variables employment and turnover belong to the concept family denoting economic performance of businesses.

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the noun. This may typically be an abbreviation of its name or a systematic number.

Characteristic	Description
Title	The official multilingual name for the concept family.
Description	Multilingual description of the concept family, including its purpose, its main subject areas etc.
Conceptual variables	A list of conceptual variables associated with the concept family. [--> Conceptual variable]
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered Item]

2.3.1.2 Conceptual variable

The conceptual variable defines the general concept of a variable (e.g. *income*, *employment*), independent of its relation to a statistical unit type or its use in a statistical activity. Thus, a conceptual variable is not a variable itself, but a concept from which a variable is derived.

More special definitions can be provided when combining a conceptual variable with a statistical unit type making up a fixed characteristic or an object variable. Conceptual variables can be grouped in concept families.

Example:

The conceptual variables employment, turnover, compensation of employees and investments belong to the concept family denoting economic performance of businesses.

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the noun. This may typically be an abbreviation of its name or a systematic number.
Title	The official multilingual name for the conceptual variable.
Description	General multilingual description of the conceptual variable, including its purpose, its main subject areas etc.
Families	The conceptual variable may belong to several concept families. [--> Concept family]
Stat. unit types	A number of statistical unit types may be associated with the conceptual variable (via fixed characteristics or object variables). [--> Statistical unit type]
Characteristics	A number of statistical characteristics (fixed characteristics and object variables) may be based on and refer to the conceptual variable. [--> Statistical characteristic]
Related concepts	The conceptual variable may be related to several similar concepts or concepts referring to the same social economic phenomena. [--> Conceptual variable]
Footnotes	A number of footnotes or remarks may be defined for the conceptual variable. [--> Footnote]

Characteristic	Description
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.3.1.3 Statistical characteristic

Statistical characteristics are defined as characteristics of a statistical unit type. A statistical characteristic can either be fixed (fixed characteristic) or variable (object variable). If the characteristic delimits a statistical unit type, it is a fixed characteristic (which cannot vary). Object variables refer to variable characteristics of a certain statistical unit type. Statistical characteristics are inherited by all sub-types of the statistical unit type.

Examples:

- *Applying the conceptual variable of employment to the STATISTICAL CHARACTERISTIC having employees of the STATISTICAL UNIT TYPE establishment generates the OBJECT VARIABLE number of employees of establishments.*
- *For the statistical unit type employer, number of employees is an object variable, because the various instances of the type can have different numbers of employees.*
- *51.874 is the SIC- category that, applied to the object variable kind of activity of the statistical unit type establishment, results in defining the subtype SIC 51.874 establishment*

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the statistical characteristic. The identifier should be unique among all characteristics of a statistical unit type and all its sub-types. The identifier should contain identification for the statistical unit type it belongs to, and the conceptual variable it is based on.
Title	The official multilingual title for the statistical characteristic.
Description	General multilingual description of the statistical characteristic, including its purpose etc.
Stat. unit type	The statistical characteristic is a characteristic of a certain statistical unit type. The statistical characteristic exists only in the context of this statistical unit type or its sub-types. [--> Statistical unit type]
Supertype characteristics	The statistical characteristic may be associated with other statistical characteristics defined for supertypes of the statistical unit. <i>Number of employees for a trading establishment may be associated with number of employees for an establishment.</i> [--> Statistical characteristic]
Subtype characteristics	The statistical characteristic may be associated with other statistical characteristics defined for subtypes of the statistical unit. <i>Number of employees for an establishment may be associated with number of employees for a trading establishment.</i> [--> Statistical characteristic]

Characteristic	Description
Conceptual variable	The conceptual variable to which the statistical characteristic is related. This provides a general description of the meaning of the statistical characteristic without explicit reference to any particular statistical unit type. <i>Employment, turnover.</i> [--> Conceptual variable]
Conceptual domain	The conceptual domain for the statistical characteristic. [--> Conceptual domain]
Subject areas	A list of subject areas in which the statistical characteristic is used. [--> Subject area]
Footnotes	A list of footnotes or remarks may be defined for the statistical characteristic. [--> Footnote]
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.3.1.4 Statistical unit type

A statistical unit type describes a class of statistical objects and its relation to supertype(s) - generalisations - and subtypes - specialisations. Thus, applying (mapping) a category of a conceptual value domain on a statistical unit type, defines a subtype by intension and a class of objects by extension.

Example:

Applying (mapping) the category female of the characteristic gender to the statistical unit type person, defines the subtype woman by intension and the class of all women by extension.

By doing so, a variable characteristic of a statistical unit type, may turn into a fixed characteristic of its subtype.

Example:

For a person, gender is a variable characteristic; for a woman, gender is a fixed characteristic.

Notice the difference between:

- A category of a classification and a statistical unit (sub)type: categories are “statistical unit neutral”. E.g., we can map the category *female* on persons, but also on nouns. Or, we can map the category “65 or more years old” of an age classification both on persons, trees and cars.
- A category and a class (of statistical units). “65+” being an age category, we get “elderly people” when mapping it on the statistical unit type *person*, while we get “old timers” when mapping it on *car*.

Special statistical unit types are elementary unit types and aggregated unit types.

A statistical unit type collects statistical characteristics in different roles. Fixed characteristics describe delimiting characteristics, which define a statistical unit type as subclass of one of its generalisations. Object variables define all variable characteristics regardless of their role as alpha, beta or gamma variables, which are defined in the context of a cube or table, only. Specifying variables are a subset of object variables, which are used for further specialisations of the statistical unit type into sub-types.

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the statistical unit type. This may typically be an abbreviation of its name or a systematic number.
Title	The official multilingual title or name of the statistical unit type.
Description	A general multilingual description of the statistical unit, including its purpose, main subject areas etc.
Type	The type of a statistical unit type describes whether the statistical unit type belongs to a <p style="margin-left: 40px;">cube unit type (aggregated or elementary unit)</p> <p>or to a</p> <p style="margin-left: 40px;">register unit type (elementary unit)</p>
Characteristics	<p>Statistical characteristics describe the characteristics of a statistical unit type. Within a statistical unit type, statistical characteristics are either fixed characteristics (i.e. values of delimiting variables) or (specifying) object variables, which might be mandatory or optional. Characteristics of a statistical unit type are the variables represented in cubes or registers belonging to the statistical unit type.</p> <p><i>For the statistical activity structural statistics on wholesale and retail trade:</i></p> <ul style="list-style-type: none"> - <i>Kind of activity and region are delimiting characteristics of the statistical unit type trade establishment, as the population of establishments is confined to Norway and traders;</i> - <i>However, kind of activity is a specifying characteristic as well, as the population is further subdivided according to lower levels of the industrial classification.</i> <p>[--> Statistical characteristic]</p>
Conceptual variables	<p>This is a list of conceptual variables associated with the statistical unit type.</p> <p>[--> Conceptual variable]</p>
Object variables	<p>The list of referenced object variables contains all object variables defined for the statistical unit type in the context of different statistical surveys or activities. Object variables are those statistical characteristics, which are not fixed. Object variables define all variable characteristics regardless of their role as beta or gamma variables.</p> <p>[--> Object variable]</p>

Characteristic	Description
Specifying variables	<p>A specifying variable is an object variable (classifying cube variable), which defines the categories, that allow further specifications (break down) of the statistical unit type into sub-types and corresponding populations/classes into sub-populations/classes. Usually, specifying variables are gamma variables. Specifying variables are a subset of statistical characteristics.</p> <p><i>Kind of activity is a specifying variable in the statistical activity structural statistics on wholesale and retail trade as the population of wholesale and retail establishments is subdivided according to categories of the industrial classification (NACE). In other words: the statistical unit type wholesale/retail trader is specified into sub-types.</i></p> <p>[--> Object variable]</p>
Measures	<p>A measure is a quantitative object variable, which is usually an optional object variable. Measures can be beta (quantifying) or gamma (classifying) variables. Measures are a subset of statistical characteristics.</p> <p><i>Number of employees (of establishments) in the trade sector is a measure.</i></p> <p>[--> Object variable]</p>
Delimiting characteristics	<p>A delimiting characteristic is a classifying characteristic with a specific category value, which specialises a statistical unit type to a more specific sub-type (e.g. <i>person to male</i>). Specialisation of a statistical unit type is strongly related to one or more delimiting characteristics, which contain the categories that determine the specialisation (e.g. <i>sex for person</i>).</p> <p>Delimiting characteristics are a subset of statistical characteristics of the statistical unit type.</p> <p><i>In the statistical activity structural statistics on wholesale and retail trade, region is a delimiting characteristic, because the category Norway is used to confine the population to Norwegian establishments.</i></p> <p>[--> Fixed characteristic]</p>
Subtypes	<p>A subtype is a statistical unit type, which is a specialisation of another statistical unit type. Usually, a subtype defines additional characteristics, turns over characteristics and specifying variables and turns one or more object variables into delimiting characteristics. Subtypes are often referenced in terms of "IS A", e.g. <i>a woman is a person</i>.</p> <p>Subtypes will inherit fixed characteristics from the super type and all object variables, which not delimit the subtype. Subtypes will create subclasses for all classes associated with its supertype.</p> <p>Variables inherited to subtypes might narrow the conceptual domain for the variable.</p> <p><i>Trading establishment is a subtype of the supertype establishment.</i></p> <p>[--> Statistical unit type]</p>
Supertypes	<p>A supertype is a statistical unit type, which is a generalisation of another statistical unit type. Going from a subtype to a supertype, means promoting a fixed or delimiting characteristic to a specifying variable or object variable.</p> <p><i>Establishment is a supertype of trading establishment</i></p> <p>[--> Statistical unit type]</p>

Characteristic	Description
Component types	<p>A component type describes objects as part of another object. A statistical unit type may have any number of component types. Object variables of component types (component variables) may have different relationships to corresponding object variables in the composite object (composite variables).</p> <p>Component-types are often referenced in terms of 'HAS ', e.g. <i>a household has members; an enterprise has establishments.</i></p> <p>1) The component variable sometimes inherits from the composite variable e.g. <i>ownership.</i></p> <p>2) The composite variable is sometimes the aggregate of the component variables</p> <p><i>The number of employees of an enterprise is the sum of the number of employees of its establishments.</i></p> <p>3) Sometimes there is no relationship between composite and component variable e.g. <i>colour.</i></p> <p>[--> Statistical unit type]</p>
Matrices	<p>A number of cubes can be associated with a statistical unit type. Each cube comprises variables appearing in a certain context (e.g. <i>in a survey or table</i>).</p> <p>[--> Matrix]</p>
Populations	<p>A number of populations for the statistical unit type may exist in statistical databases or might be defined.</p> <p>[--> Population]</p>
Subject area	<p>A statistical unit type is defined from the perspective of a subject area, which is specific for this subject area. Especially subtype/supertype relations are defined from the particular view of the subject area.</p> <p>[--> Subject area]</p>
Footnotes	<p>A number of footnotes or remarks may be defined for the statistical unit type.</p> <p>[--> Footnote]</p>
Registration	<p>The registration provides administrative information in accordance with ISO/IEC 11179.</p> <p>[--> Registered item]</p>

2.3.1.5 Fixed characteristic

A fixed characteristic is a statistical characteristic that describes an essential characteristic of a statistical unit type and is part of its definition. A fixed characteristic refers to exactly one category of the conceptual domain associated with the statistical characteristic, as opposed to an object variable, which is a statistical characteristic, but may refer to any category of the associated conceptual domain.

Typically, fixed characteristics (e.g. *male in the definition of male person*) result from applying a category (male) of an object variable (sex of a person) while creating a sub type of a statistical unit type or a sub population, from a given statistical unit type or population. There are, however, many other fixed characteristics, which are often not explicitly expressed, but which allow distinguishing objects belonging to different statistical unit types (as enterprises or persons, which can be considered both as social economic phenomena).

Example:

51.874 is the NACE category that applied to the object variable kind of activity of the statistical unit type establishment, results in defining the subtype NACE 51.874 establishment

Characteristic	Description
Stat. characteristic	The fixed characteristic is a statistical characteristic. The statistical characteristic describes the general attributes for the fixed characteristic. [--> Statistical characteristic]
Category	The category must correspond to exactly one of the categories of the conceptual domain, the fixed characteristic is related to. <i>51.874 is a category in NACE</i> All statistical units belonging to the statistical unit subtype defined by fixing a category of the statistical characteristic have by definition this category as value for this statistical characteristic.

2.3.1.6 Object variable

An object variable defines the concept of a variable in connection with a defined statistical unit type (*e.g. the income of a person*) and a conceptual domain. Usually the meaning of a variable is defined in connection with a statistical unit type, only (*e.g. size may have different meanings for person and car*).

Object variables are related to conceptual domains, which conceptually define a set of categories or a measurement unit type (*e.g. weight or currency*).

An object variable is a statistical characteristic. This generalization of the object variable can be inherited by all sub-types of the related statistical unit type, while the specialization can change from a variable to a fixed characteristic. The conceptual definition of the variable, the object-neutral concept of a variable, is described as conceptual variable, which the object variable refers to.

For wholesale trading (NACE 51)-establishment, kind of activity is an object variable, as the NACE allows for further subdivision of 51.

Specialization of a 51-establishment to a 51.874 establishment changes the object variable kind of activity into a fixed characteristic.

Synonym: Variable Characteristic

Characteristic	Description
Stat. characteristic	An object variable is a statistical characteristic and inherits all the characteristics of the statistical characteristic. The statistical characteristic describes the general attributes of the object variable. [--> Statistical characteristic]

Characteristic	Description
Mandatory	<p>Indicates whether the object variable is mandatory.</p> <p>An object variable, which refers to a necessary characteristic of a statistical unit type, is considered as mandatory.</p> <p>Necessary characteristics of a statistical unit type are those, which have a value for all objects of the related type.</p> <p>For an establishment, <i>kind of activity</i> is a mandatory object variable, because units without economic activity do not comply with the definition of establishment;</p> <p>For an establishment, <i>number of employees</i> is not a mandatory object variable, because <i>having employees</i> is not a necessary characteristic of establishments</p>
Component variables	<p>A composite variable is a variable that is created based on one or more component variables. If the object variable is composite, then the component variables may be listed here. Composite variable values are usually derived from the component variable values, <i>e.g. by calculating the sum or a percentage.</i></p> <p><i>Number of employees in an establishment (component variable) is a component of the number of employees in an enterprise (composite variable).</i></p> <p>[--> Object variable]</p>
Composite variables	<p>If the object variable is a component of one or more composite object variables, then the composite variables may be listed here.</p> <p><i>Number of employees in an enterprise (composite variable) is the sum of the number of employees in its establishments (component variables).</i></p> <p>[--> Object variable]</p>
Contextual variables	<p>A number of contextual variables may be associated with the object variable.</p> <p>The contextual variable describes specific operationalisations of the object variable used in a certain statistical activity. It refers also to more operational metadata such as value domain or derivation rules.</p> <p>[--> Contextual variable]</p>

2.3.2 Contextual variables

Contextual variables describe variables in the context of a statistical activity. Typically, contextual variables belong to registers or cubes. Depending on the way in which contextual variables are used there can be different sub types. Subtypes for contextual variables also express different roles of variables in different contexts (e.g. *classifying variable and categorical variable in cube and register respectively*).

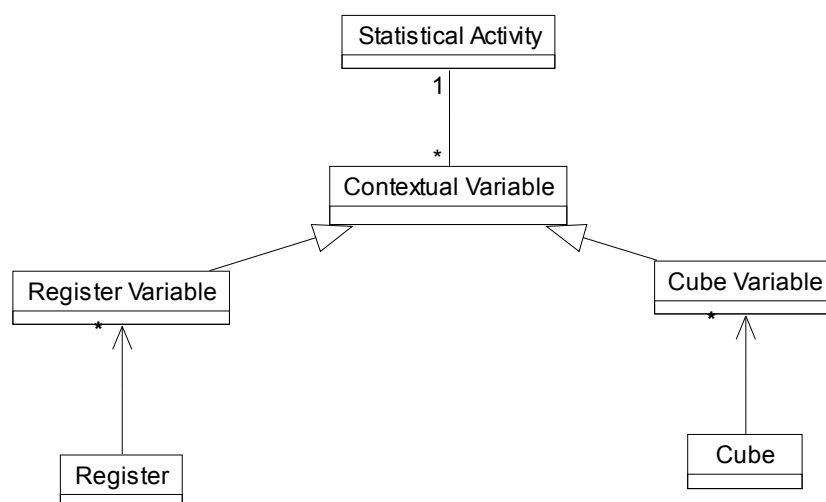


Figure 6 Contextual variable

2.3.2.1 Contextual variable

A contextual variable defines a variable in the context of a statistical activity. Contextual variables refer to statistical characteristics (object variables), which provide a standard definition for the variable. Within the context of a statistical activity, the statistical characteristic becomes more specific, which is described in the contextual variable.

While an object variable is associated with a statistical unit type and not with a specific statistical activity, a contextual variable refers to one or more matrices (registers or cubes), in which it appears.

The contextual variable is associated with the same statistical unit type as the object variable. The value domain must be consistent with the conceptual domain of the object variable, i.e. it must refer to the same conceptual domain as the related object variable.

Contextual variables can be defined as register variables or cube variables.

Example:

Number of employees is a register variable in the Norwegian register of establishments. Kind of activity is an example of a cube variable.

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the context variable. This may typically be an abbreviation of its name or a systematic number. The identifier is unique within the scope of a statistical activity.
Title	The official multilingual title of the contextual variable.

Characteristic	Description
Description	Detailed multilingual description of the contextual variable, including its use in the context of its associated object variable or statistical activity. This should describe the conceptual background of the variable, i.e. the motivation for creating the variable, its purpose and the conceptual meaning to the degree that this is not defined in the description of the associated object variable.
Factor	Factor with which the measurement unit is multiplied, e.g. <i>1 000 for the measurement unit 1 kg, when the observations are measured in tons.</i>
Component variables	<p>A composite variable is a variable that is created based on one or more component variables. If the contextual variable is composite, then the component variables may be listed here. Composite variable values are usually derived from the component variable values, e.g. <i>by calculating the sum or a percentage.</i></p> <p><i>Wages and salaries, interest and profits are component-variables of the object variable income of a person</i></p> <p>[--> Contextual variable]</p>
Composite variables	<p>If the contextual variable is a component of one or more composite variables, then the composite variables may be listed here.</p> <p><i>Following the preceding example, income can be seen as a composite variable.</i></p> <p>[--> Contextual variable]</p>
Value domain	<p>An enumerated or numerical value domain that describes permissible values (or categories) is associated with the contextual variable.</p> <p>The value domain can be associated with a measurement unit e.g. <i>the value domain for the object variable income of Norwegians can be associated with the measurement unit NOK.</i></p> <p>[--> Value domain]</p>
Object variable	<p>The object variable associated with the contextual variable.</p> <p>The object variable is associated with a conceptual domain. The latter can be associated with a measurement unit type.</p> <p><i>For wholesale trading (NACE 51)-establishment, kind of activity is an object variable, as the NACE allows for further subdivisions of 51.</i></p> <p>[--> Object variable]</p>
Activity	<p>The statistical activity in which the contextual variable has been defined.</p> <p><i>The contextual variable “number of employees” (in the Norwegian trade sector) is defined in the context of the statistical activity “The structural statistics on wholesale and retail trade”</i></p> <p>[--> Statistical activity]</p>

Characteristic	Description
Matrices	<p>A number of matrices associated with the contextual variable. A matrix can be a register or a cube.</p> <p><i>Final Observation Register for the Structural Surveys of Statistics Norway for the financial year</i></p> <p>[--> Matrix]</p>
Collection elements	<p>A number of data collection elements associated with the contextual variable.</p> <p><i>Estimation: Employment figures for the establishments are mainly obtained from the Register of Employees and Employers, but some are estimated on the basis of the establishments wage costs or turnover. The Employment figures are revised, based on information from other sources, before dissemination.</i></p> <p>[--> Data collection element]</p>
Footnotes	<p>A number of footnotes or remarks may be defined for the contextual variable.</p> <p>[--> Footnote]</p>
Registration	<p>The registration provides administrative information in accordance with ISO/IEC 11179.</p> <p>[--> Registered item]</p>

2.3.2.2 Register variable

Register variables are contextual variables that belong to a register. Register variables can be categorical or numerical variables. Some register variables can be identifier or link variables.

Examples;

Number of employees and kind of activity are register variables for the Norwegian register of establishments.

Characteristic	Description
Contextual variable	<p>A register variable is a specific contextual variable. It inherits all the characteristics of the contextual variable <i>e.g. the matrices of the contextual variable contain a number of registers in which the register variable is used.</i></p> <p>[--> Contextual variable]</p>
Variable type	<p>Variable types for registers are:</p> <ul style="list-style-type: none"> numerical - A numerical variable is a register variable that assumes a quantitative value. <p style="text-align: center;"><i>Number of employees is a numerical variable in the Norwegian register of establishments;</i></p> categorical - A categorical variable is a register variable that assumes a qualitative value (<i>e.g. a category of a classification</i>). <p style="text-align: center;"><i>Kind of activity is a categorical variable in the Norwegian register of establishments.</i></p>

Characteristic	Description
Related unit type	A register may contain register variables that refer to other statistical unit types than the register unit type. [--> Statistical unit type]

2.3.2.3 Cube variable

Cube variables are contextual variables that belong to a cube. Cube variables are either quantifying or classifying variables depending on their variable type.

Examples;

- *Kind of activity, Number of employees*

Characteristic	Description
Contextual variable	A cube variable is a specific contextual variable. It inherits all the characteristics of the contextual variable <i>e.g. the matrices of the contextual variable contains a number of cubes in which the cube variable is used.</i> [--> Contextual variable]
Variable type	Variable types for cubes are: Classifying - A classifying variable is a cube variable that usually acts as a dimension in a cube. Classifying variables are ultimately derived either from numerical variables transformed into groups or from qualitative variables. Classifying variables can be derived from other classifying or quantifying variables. <i>If, in a cube, the total population of establishments is subdivided into sub populations according to size-classes, the numerical register variable number of employees acts as a classifying cube variable.</i> Quantifying - A quantifying variable is a cube variable that assumes a numerical value, which is associated with a measurement unit. Quantifying variables are (ultimately) always derived from quantitative variables by aggregating, counting, calculating the mean or other statistical functions. A quantifying variable can be derived from another quantifying variable. <i>If, in a cube, the total number of employees for a population of establishments is recorded, the numerical register variable number of employees acts as a quantifying cube variable.</i>

2.3.3 Value domains and measurement units

Value domains and measurement units play a role for context related as well as for context independent variables. At the contextual level, the definition of value domains in relation with contextual variables is precise and operation oriented: at this stage of the definition process, the codes for enumerated value domains as stored in the data files are known. At the conceptual level, more

meaning oriented the value domains are defined by means of categories (rather than codes) and measurement unit types (rather than measurement units).

For a value domain at the conceptual level, the term of conceptual domain will be used, whereas the term value domain will be used at the contextual level.

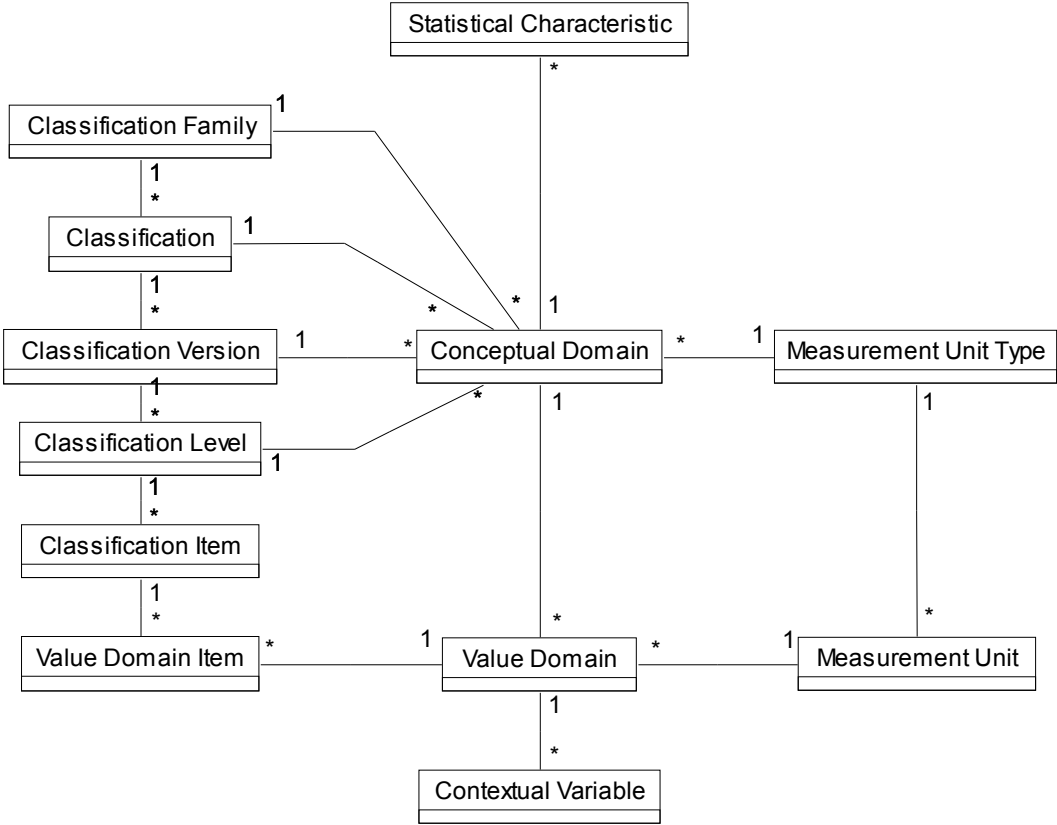


Figure 7 Value domains and measurement units

2.3.3.1 Conceptual domain

Conceptual domains can be enumerated (listed) or non-enumerated (not listed) i.e. the conceptual domain defines a set of categories or includes a declaration of the type of measurement unit that can be used for certain statistical characteristics (object variables). Categories are generally defined in classifications. In that case, the conceptual domain may define different degrees of detail referring to classifications, classification versions or levels of classifications.

Examples:

Standard for industrial classifications (SIC)

Probability

Characteristic	Description
----------------	-------------

Identifier	A unique, language independent identifier is used to identify the conceptual domain. This may typically be an abbreviation of its title or a systematic number. The identifier has to be unique among all conceptual domains.
-------------------	---

Title	The title is a short multilingual label for the conceptual domain. It could include the characteristic and the purpose of the conceptual domain.
--------------	--

Characteristic	Description
Description	General multilingual description of the conceptual domain, including a description of its conceptual background and its contextual implementations.
Classification family	When the conceptual domain is enumerated it may refer to a classification family, which refers to a number of classifications containing possible categories for the value domain. [--> Classification family]
Classification	When the conceptual domain is enumerated it may refer to a classification. <i>The conceptual domain for the object variable kind of activity of establishments is enumerated and may refer to NACE</i> [--> Classification]
Classification version	When the conceptual domain is enumerated it may refer to a classification version. In this case, it must also refer to the classification the classification version belongs to. <i>The classification version currently applying for the classification NACE is NACE rev.3</i> [--> Classification version]
Classification level	When the conceptual domain is enumerated it may refer to a classification level. In this case, it must also refer to the classification version the classification level belongs to. <i>NACE rev. 3 has 5 levels, the third of which is the division level</i> [--> Classification level]
Classification items	When the conceptual domain is enumerated it may refer to classification items. In this case, it must refer to the classification version and the classification level, the classification items belongs to. <i>In NACE rev.3, the division level contains 60 classification items, one of which is 50 Sale, maintenance and repair of motor vehicles and motorcycles, retail sale of automotive fuel</i> [--> Classification item]
Measurement unit type	When the conceptual domain is non-enumerated it refers to a measurement unit type. In this case, all associated contextual variables are measured in a measurement unit of the referenced measurement unit type. <i>For the object variable number of employees of establishments, number is the measurement unit type.</i> [--> Measurement unit type]
Value domains	A number of value domains based on this conceptual domain might be defined. [--> Value domain]
Statistical characteristics	A number of statistical characteristics associated with conceptual domain. [--> Statistical characteristic]

Characteristic	Description
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.3.3.2 Value domain

A value domain defines the specific valid values (domain) for a contextual variable, i.e. all the possible values that will finally be stored in the data items. The scope and the meaning of the possible values are defined within the frame of the conceptual domain that the value domain is associated with. Contextual value domains can be enumerated or non-enumerated.

Enumerated value domain:

An enumerated value domain is associated with a number of value domain items. It might also be associated with a measurement unit. These value domain items may represent

- Classification items: in most of the cases, the enumerated value domain, is then associated with a classification level via a conceptual domain. Each value domain item corresponds to a classification item, and the code of the value domain item might be identical with the code of the classification item or not, depending on operational considerations. There might even be several operational codes defined for the same category in different value domains for operational reasons. In a statistical office, harmonization of operational codes of value domains should be promoted as much as possible (same operational code for same category). Indeed, the codes of the classification items are not really values but rather part of the meaning of a category describing a hierarchical relationship between categories (as in NACE) or providing an abbreviation (as for ex. in a gender classification: 'm' for male and 'f' for female).
- Items independent of a classification version: list of value domain items can be defined without any relation to a classification version. In this case, the codes, titles, meanings, etc. are defined within the scope of the value domain items.
- Value ranges: intervals or grouping of values. Value ranges are often associated with a measurement unit.

The numerical object variable number of employees can be expressed in size classes, where *number* is the measurement unit type. .

Non-enumerated value domain:

A non-enumerated value domain represents a set of quantitative values that is often associated with a measurement unit. Non-enumerated value domains are used for continuous data.

Example:

The value domain for number of employees of establishments ranges over all non-negative integers.

Synonym: Value Domain

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the value domain. This may typically be an abbreviation of its name or a systematic number. The identifier has to be unique among all value domains.

Characteristic	Description
Title	The multilingual title is a single word or group of words that designates the value domain. It is either a systematic name according to the naming conventions in the organization or a natural language name that reflects the idea or concept of this value domain.
Description	General multilingual description of the value domain, including its purpose and usage for different contextual variables.
Constraint	Constraints are used to define a set of permissible values for the domain. For enumerated value domains, this may describe the code structure for values in the domain (<i>e.g.</i> 99.99). For non-enumerated value domains this may define a pattern as well, but also an interval of numeric values. Since there is no standard for defining such constraints, the details for defining value constraints depend on agreements in the organisation.
Positions	Maximum number of significant characters or figures for any value in the value domain.
Enumerated	Indicates whether the value domain is enumerated.
Number of values	In the case of an enumerated value domain, number of categories or values available in this value domain.
Values	In the case of an enumerated value domain, this collection contains the permissible values or value domain items defined for the value domain. [--> Value domain item]
Conceptual domain	The value domain usually refers to a conceptual domain. When the conceptual domain defines a classification version and/or a level the value domain must be consistent with the definition in the conceptual domain. [--> Conceptual domain]
Measurement unit	Non-enumerated or enumerated value domains may refer to a measurement unit, which defines the unit in which the data is measured. Enumerated value domains may refer to a measurement unit, <i>e.g.</i> for value ranges (<i>grouping numerical values</i>). In this case the interval described by each value domain item is associated with the measurement unit referenced here. Non-enumerated value domains are often associated with a measurement unit. [--> Measurement unit]
Base value domain	If the value domain is a recoding of a base value domain, then the base value domain is referred to here. <i>1 letter code for sex: 'm' for male and 'f' for female - {m,f}</i> [--> Value domain]
Recodings	If the value domain is a base value domain, then this is a list of value domains that define recodes for the value domain. <i>1 digit code for sex, version 1: '1' for male and '2' for female - {1,2}</i> <i>1 digit code for sex, version 2: '0' for male and '1' for female - {0,1}</i> [--> Value domain]

Characteristic	Description
Contextual variables	A number of contextual variables associated with value domain. [--> Contextual variable]
Successor	When there are changes in a value domain the successor refers to the new value domain. [--> Value Domain]
Predecessor	When there are changes in a value domain the predecessor refers to the previous value domain. [--> Value domain]
Changes from predecessor	Describes the changes that have occurred from the predecessor value domain to the current value domain.
Footnotes	A number of footnotes or remarks may be defined for the value domain. [--> Footnote]
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.3.3.3 Value domain item

A value domain item describes a permissible value in the enumerated value domain where it is defined. The data items at the data level will use this value. There are three kinds of value domain items:

- Value domain item associated with an item of a classification version: this value domain item, in most cases, has the same code as the classification item. Nevertheless, for operational reasons, it need not have the same value or code as the classification item. Indeed the code of a classification item is not really a value but rather part of the meaning of a category describing a hierarchical relationship between categories (as in NACE) or providing an abbreviation (as in a gender classification: 'm' for male and 'f' for female).
- Value domain item independent of any classification item: in this case, its code and its title, meaning, etc. are defined within the scope of the value domain item. Independent value domain items can often be found when there are lists of values that are specific to a certain statistical activity, without any normative character: for ex. for the value domain type of heating, the following value set items could be defined: gas, petrol, solar, wood, etc. Special values such as no response, wrong response, etc. are also independent of any classification version item and are often added in value domains based on a classification version.
- Value domain item representing a value range: a value range is an interval or grouping of numerical values with a lower and an upper boundary, which can be open or closed *e.g.* (0, 110], [1, 20).

Synonym: Contextual Value Domain Item

Characteristic	Description
Code	<p>A value domain item is identified by an alphabetical, numerical or alphanumeric code, which is in line with the code structure (constraint) of the value domain. The code is unique within the value domain to which the value domain item belongs.</p> <p>For classification based value domains, the code of a value domain item is usually the code from the classification item in the originating classification version. It may, however, differ from the code in the classification version to provide a transformation from the code stored in the classification to the code stored in the data.</p> <p>For a value domain independent of a classification version, the code of a value domain item is defined within the scope of the value domain item itself.</p> <p><i>“50” is the code for the value domain item "Sale, maintenance and repair of motor vehicles and motorcycles, retail sale of automotive fuel" in NACE.</i></p>
Title	<p>The title is a multilingual label for the value domain item that describes the content of the category. When the value domain item is based on a classification item, the title is usually inherited from the classification item.</p>
Alternative titles	<p>A value domain item can be expressed in terms of one or several alternative multilingual titles. Each alternative title is associated with a title type. When the value domain item is based on a classification item, the title type is usually inherited from the classification item.</p> <p><i>Examples of title type: Short titles; Medium titles; Titles in plural form (e.g. Men, Women) for dissemination purposes; gender related titles.</i></p>
Description	<p>General multilingual description of the value domain item, including a more detailed explanation of the value meaning, borderline definitions and the way of using it.</p> <p>The Description field in Value Domain Item needs to have a reference to the Conceptual Value Domain from which the description comes. The description must correspond to a category contained in a CVD and correspond to a category in a Classification Scheme.</p>
Lowest value	<p>When a value domain item represents a value range or interval of numerical values, there must be a lower boundary for this value range.</p> <p><i>For the value domain of age groups, the age group 1 to 5 has a lower boundary of 1.</i></p>
Highest value	<p>When a value domain item represents a value range or interval of numerical values, there must be an upper boundary for this value range.</p> <p><i>For the value domain of age groups, the age group 1 to 5 has an upper boundary of 5.</i></p>
Special value	<p>Indicates whether or not the value domain item is a special value.</p> <p><i>Missing value.</i></p>
Value domain	<p>This is the value domain the item belongs to.</p> <p>[--> Value domain]</p>

Characteristic	Description
Classification item	When the value domain item refers to a classification item, the classification item is indicated here. However, the value domain can include value domain items that are not present in the classification <i>e.g. no response</i> . [--> Classification item]
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.3.3.4 Measurement unit type

The measurement unit type defines the type of a measure *e.g. mass or currency*. The measurement unit type groups all measurement units, which can be converted into each other. Each measurement unit type has a standard measurement unit that is used for conversion between different measurement units (*e.g. kilogram for mass*).

Example:

For the object variable number of employees of establishments, number is the measurement unit type

Synonym: Dimensionality

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the measurement unit type. This identifier has to be unique within all measurement unit types.
Title	The official multilingual title of the measurement unit type.
Description	Short multilingual description of the measurement unit type.
Conceptual value domains	A list of non-enumerated conceptual domains that refer to this measurement unit type. [--> Conceptual domain]
Measurement units	A list of measurement units defined for this measurement unit type. <i>For the object variable "turnover of trade establishments" currency may be the measurement unit type, with Norwegian Crown as the measurement unit.</i> [--> Measurement unit]
Standard unit	Each measurement unit type refers to one of its related measurement units as a standard measurement unit, which is used for conversion between different measurement units (<i>e.g. kilogram for mass</i>). All measurement units of the same measurement unit type can be expressed by a factor relative to the standard measurement unit. [--> Measurement unit]
Footnotes	A number of footnotes or remarks may be defined for the measurement unit type. [--> Footnote]

Characteristic	Description
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.3.3.5 Measurement unit

A measurement unit is the metric for a measurement in terms of an official unit of measurement (gram, dollar, meter) and a factor with which the unit is multiplied (*e.g. in 1000 grams, 1000 dollars or 1000 meters*). Measurement units can be based on different measurement unit types such as weight, height, currency, duration etc. Measurement units can be transformed into one another as long as they refer to the same measurement type.

Characteristic	Description
Identifier	This is a unique and language independent identifier for the measurement unit, which may typically be the official abbreviation of its name (<i>e.g. 1000 m</i>). The identifier has to be unique within the scope of all measurement units.
Title	This is the official multilingual name of the measurement unit. <i>Kilogram, ton.</i>
Abbreviation	This is the official multilingual abbreviation of the measurement unit. <i>kg</i>
Description	Short multilingual description of the measurement unit.
Factor	Factor with which the standard measurement unit is multiplied. <i>1 000 for the measurement unit ton, if the standard measurement unit is kg.</i>
Type	Each measurement unit should refer to a measurement unit type. [--> Measurement unit type]
Standard unit	A standard measurement unit is used for conversion between different measurement units. All measurement units of the same measurement unit type can be expressed by a factor relative to the standard measurement unit. [--> Measurement unit] <i>Example: kilogram for measurement unit type mass.</i>
Value domains	A list of value domains associated with the measurement unit. [--> Value domain]
Footnotes	A number of footnotes or remarks may be defined for the measurement unit. [--> Footnote]
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.4 Data structure

The content of a survey or sample, i.e. the variables, is described separately in a matrix definition as register or cube. Thus, a matrix definition expresses a specific statistical view to social economic phenomena. A matrix is the generalisation of register or cube and contains a list of contextual variables, for which data has been investigated in the survey or sample. This variable list mainly defines the data structure (or the content) of a statistical survey, sample or cube. Usually, a technical counterpart (technical data structure) is required, which describes the data format and position in a data record. This, however, is not part of the considerations in this paper.

2.4.1 Unit types, registers, cubes and tables

Statistical unit types describe classes of objects that statistics relate to in different phases of statistical production. Data collected in registers or cubes usually relate to specific statistical unit types. Depending on the processing phase in statistical production different kinds of statistical unit types are considered (as register unit types in the input phase or cube unit types on the output side).

Registers and cubes are basic metadata object types referring to contextual variables of different types. Registers and cubes define the intension of a data collection in terms of variable definitions, which are provided with the register or cube. Registers also define par excellence the extension of (elementary) units, while cubes define the extension of (aggregate) units. Thus, registers and cubes are the key objects for describing data on a conceptual level.

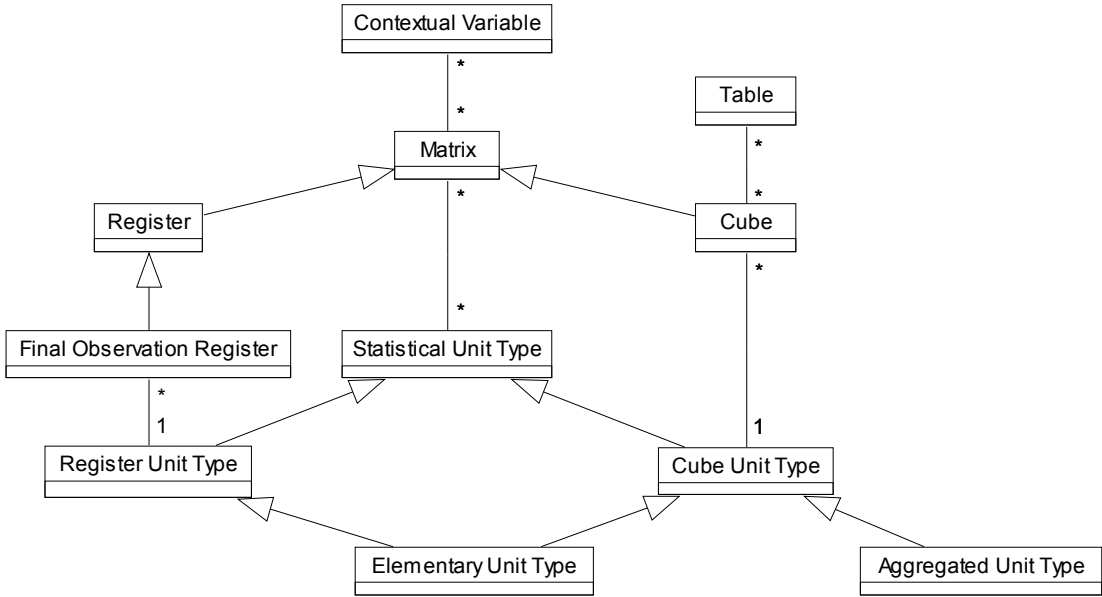


Figure 8 Registers and Cubes

The above figure shows the relations between statistical unit type and related matrix object types.

2.4.1.1 Register unit type

A register unit type describes an elementary unit type, which is related to data collected in final observation registers, by individuating and identifying the instances of the unit type.

Characteristic	Description
Unit type	<p>A register unit type is a statistical unit type and inherits all the characteristics of the statistical unit type.</p> <p><i>The statistical activity structural statistics of wholesale and retail trade of Statistics Norway uses a REGISTER containing establishments. Ericsson AS Grimstad is one of them.</i></p> <p>[--> Statistical unit type]</p>
Registers	<p>A list of final observation register referring to this register unit type.</p> <p>[--> Final observation register]</p>

2.4.1.2 Cube unit type

A cube unit type is an elementary or aggregated unit type, which describes statistical characteristics represented in cubes.

Characteristic	Description
Unit type	<p>A cube unit type is a statistical unit type and inherits all the characteristics of the statistical unit type. It is the base for one or more cubes.</p> <p><i>The statistical activity “structural statistics of wholesale and retail trade” of Statistics Norway generates cubes, in which the trade establishment is the elementary unit type, and the trade sector is the aggregated unit type.</i></p> <p>[--> Statistical unit type]</p>
Cubes	<p>A list of cubes referring to this cube unit type.</p> <p>[--> Cube]</p>

2.4.1.3 Aggregated unit type

Aggregated unit types identify data on aggregated levels.

Characteristic	Description
Unit type	<p>Aggregated unit types are cube unit types and inherit all the characteristics of these.</p> <ul style="list-style-type: none"> - <i>Females in Sweden from 30 to 34 years old.</i> - <i>The Norwegian trade sector</i> <p>[--> Cube unit type]</p>

2.4.1.4 Elementary unit type

An elementary unit type is a statistical unit type for whose instances information is sought and for which statistics are ultimately compiled. Elementary unit types (*e.g. person, establishment or events such as birth or accident*) refer to elementary objects (*e.g. Ericsson AS Grimstad*). An elementary unit type describes the type of objects defined in an initial or final observation register or presented in a cube. Thus, elementary unit types may appear as statistical register or cube units.

Characteristic	Description
Register unit	An elementary unit type can be a register unit type. <i>In the Norwegian Final Observation Register: the establishment</i> [--> Register unit type]
Cube unit	An elementary unit type can be a cube unit type. <i>In the statistical activity “structural statistics of wholesale and retail trade” of Statistics Norway, the trade establishment is the elementary cube unit type.</i> [--> Cube unit type]

2.4.1.5 Collection unit type

The collection unit type refers to the type of objects during the input (or collection) phase of a statistical process. These can be different from register unit types, which reflect the perspective of statistical survey. Collection unit types are the unit types about which data are supplied (often materialized in questionnaires).

Example:

For the statistical activity “structural statistics of wholesale and retail trade” of Statistics Norway, the collection unit type is the same as the register unit type i.e. establishment.

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the collection unit type. This may typically be an abbreviation of its title or a systematic number.
Title	The official multilingual title of the collection unit type.
Description	General multilingual description of the collection unit type, including its purpose, definition etc.
Initial observation registers	A list of initial observations registers referring to this collection unit type. [--> Initial observation register]
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.4.1.6 Matrix

A matrix describes the content of one or more data collections on micro or macro data level (register or cube), which is the intension of the data collections. It defines the included contextual variables and the statistical unit types as well as the rules for building the matrix. A matrix always instantiates in a register or cube, i.e. it never appears just as a matrix.

The matrix defines the content of a data collection by means of contextual variables, i.e. it does not refer to a specific population or data collection. The matrix does not describe the extension for data collections.

A matrix may refer to a number of data collections, which contain data for the contextual variables described in the matrix. Thus, a number of data collections with different extensions can be described by the same matrix as long as they follow the same intension.

Example:

Final Observation Register for the Structural Surveys of Statistics Norway for the financial year

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the matrix. This may typically be an abbreviation of its title or a systematic number.
Title	The title of the cube is a short multilingual characteristic of the content of the cube or register (matrix).
Description	General multilingual description of the matrix, including the content and purpose.
Reference time	The reference time is a time point or period when the phenomena, the matrix describes, has happened. The reference time may also be relative to another reference time <i>e.g. previous year</i> . Usually the reference time of the contextual variables is identical with the reference time of the matrix, but in some cases contextual variables may refer to different time points.
Contextual variables	A list of contextual variables associated with the matrix. [--> Contextual variable]
Stat. unit types	A list of statistical unit types for the matrix. [--> Statistical unit type]
Data collections	A list of data collections for the matrix. Each data collection represents the content of the data set and can be stored in different data sources. [--> Data collection]
Activity	The statistical activity in which the matrix is defined. [--> Statistical activity]
Successor	When the matrix is replaced by a successor, due to small changes in a new statistical activity instance, the successor matrix is referenced here. [--> Matrix]
Predecessor	When the matrix is a successor of another matrix, the predecessor matrix is referenced here. [--> Matrix]
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.4.1.7 Register

A register (initial or final observation register) is a matrix that describes the content of one or more collections of elementary unit types for a given statistical unit type through the definition of the

register unit type and a set of contextual variables. A register can also refer to other statistical unit types (e.g. *the enterprise a person is working for*).

Example:

Final Observation Register for the Structural Surveys of Statistics Norway for the financial year.

Characteristic	Description
Matrix	The matrix on which the register is based. A register is a matrix and inherits all the characteristics of the matrix. [--> Matrix]
Identifier variables	The values of one or more variables may identify the instances of a unit type in a register (e.g. <i>person number in a person register</i>). [--> Register variable]
Link variables	Link variables in a register are variables that refer to an observation object in another register. <i>Reference to the household in a person register.</i> [--> Register variable]
Numerical variables	Numerical variables in a register are variables that define quantitative values measured for objects of a statistical unit type. <i>Number of employee.</i> [--> Register variable]
Categorical variables	Categorical variables in a register are variables that define certain categories for statistical units of a given statistical unit type <i>Kind of activity (SIC-code).</i> [--> Register variable]
Register unit type	The register unit type for the register. [-->Register unit type]

2.4.1.8 Initial observation register

An initial observation register describes the intension of the data collected for a statistical survey during the input or data collection phase. It may refer to one or more data collections, which describes the extension of collected data for the initial observation register. Data for an initial observation register is considered as micro data describing the characteristics of elementary units.

An initial observation register consists of contextual variables, which may belong to one statistical unit type (collection unit type). The initial observation register describes the content of data collections of collection units, which are the units about which data are supplied (often materialized in questionnaires). In some cases, collection units may differ from the reporting units.

Synonym: Conceptual initial observation register

Characteristic	Description
Register	The initial observation register is a register and inherits all the characteristics of the register. [--> Register]
Reporting org.	The reporting organisation is the organisation from which data are reported (often given as addresses or contact persons)
Collection unit type	The initial observation register has a number of collection units of a given collection unit type within a given population. Collection units are the units about which data are supplied (usually materialized in questionnaires). [--> Collection unit type]

2.4.1.9 Final observation register

A final observation register describes the intension of the data collected in a statistical survey based on a target population. It may refer to one or more data collections, which describes the extension of collected data for the final observation register. Data for a final observation register is considered as micro data describing the characteristics of elementary units.

A final observation register refers to one statistical unit type. It may, however, contain contextual variables that are related to other statistical unit types, derived or even aggregated data (*e.g. the value of turn over by type of product of an enterprise is aggregated data relating to the statistical unit type "product"*).

Synonym: Conceptual final observation register

Characteristic	Description
Register	The register describes the general properties of the final observation register. [--> Register]
Unit type	The final observation register contains information about observed object instances, which belong to the same register unit type. [--> Register unit type]
Cubes	A list of cubes that are based on the final observation register. [--> Cube]

2.4.1.10 Cube

A cube is a special intensional matrix that describes the intension of aggregated data. A cube is a generalization of a multidimensional table. The table dimensions are considered as cube dimensions. The columns in the multidimensional table are presented as observations in the cube. Usually, cubes are the basis for table generation.

Usually a cube describes the result of a matrix operation, as the aggregation of register data. It is, however, also possible to describe a cube as input for a matrix operation (*e.g. for combining aggregated data with cube data from another statistical activity*).

Cube data can be described in one or more data collections (*e.g. for different statistical activity instances or survey instances*), which contain the extensional definition of the cube.

Cube data may consist of data aggregated from a register or could also be the result of an operation on other cubes.

A cube may refer to statistical unit types in different ways:

- the elementary statistical unit type (i.e. register unit) defined by the final observation register the cube is based on
- other statistical unit types referenced in the register, either aggregated or elementary
- the aggregated unit type defined by the cube dimensions

The statistical unit type, which a cube refers to, depends on the intension of the cube. Thus, a cube with one regional dimension might refer to "Person" data as well as to "Region" data, depending on the users view of the published cube data.

Example:

Wholesale and retail trade, structural statistics: Principal figures (employment, turnover, compensation of employees, investments) by industry subclass (division) for establishments.

Characteristic	Description
Matrix	A cube is based on a matrix that defines the content and the purpose of the cube. The cube inherits all the characteristics of the matrix. [--> Matrix]
Dimensions	A list of classifying variables in the cube i.e. cube variables with variable type classifying. <i>Kind of activity and size class (employees).</i> [--> Cube variable]
Quantifying variables	A list of quantifying variables in the cube i.e. cube variables with variable type quantifying. Quantifying variables describe the measures, i.e. aggregated quantitative values, in a cube. <i>Number of establishments, number of employees; compensation of employees; turnover, etc.</i> [--> Cube variable]
Final observation registers	A cube is based on one (or more) final observation registers, which provide the data for the cube. [--> Final observation register]
Register unit	The register unit type describes the final observation register the cube is made from. If the cube is made from several final observation registers, there is usually one, which can be considered as the main register for building the cube. [--> Register unit type]
Cube unit	The cube unit type describes the perspective of the cube. It reflects the target objects that are described in the cube. <i>In the "structural statistics of wholesale and retail trade" of Statistics Norway, cubes describe the statistical characteristics of industry groups in the wholesale retail trade.</i> [--> Cube unit type]

2.4.1.11 Table

A table is an electronic or paper presentation of (usually) aggregated statistical data. A table is based on a cube that defines the lowest level of data displayed in the table. In contrast to a cube, a table might display data on different levels. It may also contain explanatory text, footnotes and other information. A table is always considered to be the result of a statistical process.

Example:

The table “principal figures by size class (number of employees) and industry subclass (division)” is one of the tables, resulting from the statistical activity “structural statistics, wholesale and retail trade” of Statistics Norway.

The unit of interest is, at the elementary level, the establishment. At the aggregate level, it is the industry group / size class.

Synonym: Conceptual table

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the table. This may typically be an abbreviation of its title or a systematic number.
Title	Official multilingual title for the table.
Description	General multilingual description of the table, including its purpose, its main subject areas etc.
Subject areas	One or more subject areas in which the table is used. [--> Subject area]
Unit of interest	The unit of interest describes the aim of the table. It reflects the target units that are reflected in the cube the table is based on. [--> Statistical unit type]
Activity	The statistical activity the table is defined in. [--> Statistical activity]
Cubes	A list of cubes the table is based on. [--> Cube]
Instances	One or more table instances can be defined for a table that refer to physical tables and the location where those tables are available. [--> Table instance]
Successor	When the table has been replaced by a successor, in a new statistical activity instance, the successor table is referenced here. [--> Table]
Predecessor	When the table is a successor of another table, the predecessor is referenced here. [--> Table]
Keywords	A number of keywords may be defined for the table. [--> Keyword]

Characteristic	Description
Footnotes	A number of footnotes or remarks may be defined for the table. [--> Footnote]
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.4.2 Populations, frame and sample

Populations denote the extension of a statistical unit type, as it appears in a statistical activity. Indeed, the notion of population is always associated with a particular statistical activity.

In contrast to registers and cubes, which describe the intension of a dataset (data collection), populations describe the extension of a dataset. Besides populations, classes of statistical unit types provide also some information about the extension of a data collection, which is, however, implicitly defined. Besides the fact, that populations may form hierarchies based on subset relationships, populations may play different roles in different phases of a statistical activity (survey, frame or target population).

The target population describes the purported extension (collection of statistical units), as defined during the design stage of a statistical activity. The target population is an ideal population, which usually differs from the survey population, which is created from a given frame.

Populations are materialized by means of frames in many cases, which provide the base for a population as register (*e.g. person or business register*).

Populations may refer to population instances, which describe the instantiation of a population in the context of a statistical activity instance (or survey instance). While the population as such describes the concept of population within a statistical activity, the population instance is the result of a process, in which the population has been developed within a statistical activity instance.

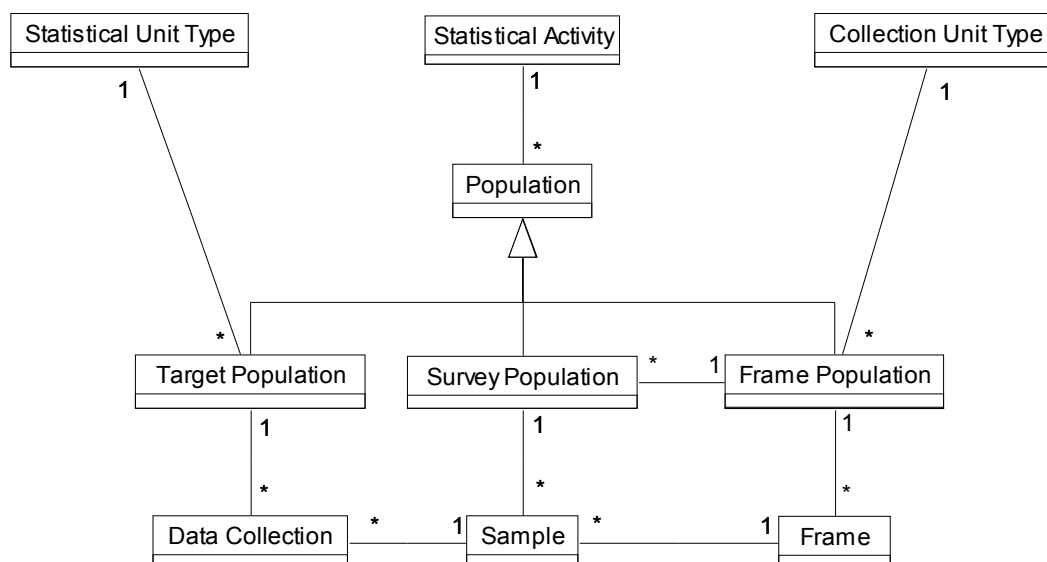


Figure 9 Population

2.4.2.1 Population

The population denotes the extension of a data collection. A population is generally delimited in space and time. It is an instantiation of a (sub) class of units of a statistical unit type, which describes the general construction rules for the population and in most cases the restrictions in time and space. In contrast to a class, the population is a particular set of statistical units.

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the population. This may typically be an abbreviation of its title or a systematic number. It is unique in the context of a statistical activity.
Title	The title is a short multilingual label for the population.
Description	Detailed multilingual description of the population, which typically includes a description of the way in which the population is defined and an estimate of the number of objects in the population.
Stat. unit type	The statistical unit type for the population. If the population is a subpopulation, then the statistical unit type is a subtype of the statistical unit type for the related frame. [--> Statistical unit type]
Activity	The statistical activity associated with the population. [--> Statistical activity]
Super populations	A list of super populations defined for this population. A super population includes all the subpopulations of the population. [--> Population]
Subpopulations	A list of subpopulations defined for this population. A subpopulation defines a subset of objects described by the population. [--> Population]
Instances	A list of population instances, which have been created for the population in different statistical activity instances. [--> Population instance]
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.4.2.2 Target population

The target population describes the purported extension (collection of statistical units, i.e. the objects of interest), as defined during the design stage of a statistical activity. The target population is an ideal population, which usually differs from the survey population, which is created from a given frame.

Characteristic	Description
Population	The target population is a population and it inherits all the characteristics of population. [--> Population]

Characteristic	Description
Target population	The target population describes the ideal population on which the statistical data resulting from a statistical activity are based [--> Population]
Frame population	When the population is a target population, it may refer to a frame population that specifies a survey population [--> Frame population]
Survey population	When the population is a target population, it may refer to a survey population, which is meant to be the best possible approximation of the target population. [--> Survey population]
Activity	The statistical activity in which the frame population is defined. [--> Statistical activity]

2.4.2.3 Frame population

The frame population is a specific population built from a frame which provides the individual collection units. The frame population describes the way to select units from the frame. The frame population specifies the survey population, but the meanings of the terms are slightly different. The survey population is the collection of the statistical units, while, in addition, the frame population also contains information about the statistical units (addresses, travel costs connected to interviews etc.) necessary to draw an optimal sample. An example of the difference between the two terms is that a statistical unit (e.g. a person) in a survey population may remain the same from one year to another, but if he/she changes his/her address, the corresponding unit in the frame population will change.

A statistical activity may refer to one or more frame populations.

Example:

- *For the STATISTICAL ACTIVITY “structural wholesale and retail trade statistics” of Statistics Norway, the FRAME POPULATION for the collection of data for the year t consists of all establishments registered as active in the Norwegian Central Register of Enterprises and Establishments in the year t, and classified as NACE-industry divisions 50, 51 or 52, including relevant information about each establishment, such as address etc.*

Characteristic	Description
Population	The frame population is a population. It inherits all the characteristics of population. [--> Population]
Frame	The frame from which the frame population is created. [--> Frame]
Target population	The target population describes the ideal population on which the statistical data resulting from a statistical activity are based [--> Population]
Activity	The statistical activity in which the frame population is defined. [--> Statistical activity]

2.4.2.4 Survey population

The survey population is the population from which the information can be obtained in the survey. It is the collection of statistical units based on the frame population and meant to be the best possible approximation of a given target population. The survey population often differs from the target population because of frame errors like overcoverage, undercoverage and wrong classifications of units.

Characteristic	Description
Population	The survey population is a population. It inherits all the characteristics of population. [--> Population]
Rules	Rules for making the sample.
Sample	The sample created from the survey population [--> Sample]
Target population	The target population describes the ideal population on which the statistical data resulting from a statistical activity are based [--> Population]
Activity	The statistical activity in which the survey population is defined. [--> Statistical activity]

2.4.2.5 Frame

The frame is the place from which the sample for a survey is taken. It is the tool or the resource that provides the individuals for a population. Thus, it describes the method for providing individuals directly or indirectly from a given source.

The frame is often based on a register (person or business register, list of phone numbers, etc.), which provides or generates the individuals for a frame population. A statistical activity may refer to one or more frames that provide the individuals for the total populations the activity is based on.

Example:

The Norwegian Central Register of Enterprises and Establishments is the FRAME for the STATISTICAL ACTIVITY "structural wholesale and retail trade statistics" of Statistics Norway:

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the frame. This may typically be an abbreviation of its title or a systematic number.
Title	The title is a short multilingual label for the frame.
Description	Detailed multilingual description of the frame, which typically describes how to use the frame or where it is available.
Frame populations	A number of frame populations may refer to the frame. [--> Frame population]
Survey populations	A number of survey populations may refer to the frame. [--> Survey population]

Characteristic	Description
Register	The frame might be stored in a register (<i>e.g. business register</i>). [--> Register]
Target population	The target population describes the purported population on which the statistical data resulting from a statistical activity are based [--> Population] <i>Example:</i> <i>For the STATISTICAL ACTIVITY “structural wholesale and retail trade statistics” of Statistics Norway, the TARGET POPULATION for the CUBES and TABLES resulting from the “wholesale and retail trade statistics in year t” is defined as: “all establishments existing at any moment in year t, and engaged in wholesale or retail trade” as their main kind of activity.</i>
Registration	The registration provides administrative information according to ISO/IEC 11179. [--> Registered item]

2.4.2.6 Sample

The sample is a subpopulation of units that has been identified and sampled from the survey population under a specified sample scheme.

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the sample. This may typically be an abbreviation of its title or a systematic number.
Title	The title is a short multilingual label for the sample.
Description	Detailed multilingual description of the sample.
Probabilities	Inclusion probabilities for each unit
Survey population	The survey population from which the sample is drawn. [--> Survey population]
Target population	The target population describes the purported population on which the statistical data resulting from a statistical activity are based [--> Population] <i>Example:</i> <i>For the STATISTICAL ACTIVITY “structural wholesale and retail trade statistics” of Statistics Norway, the TARGET POPULATION for the CUBES and TABLES resulting from the “wholesale and retail trade statistics in year t” is defined as: “all establishments existing at any moment in year t, and engaged in wholesale or retail trade” as their main kind of activity.</i>
Registration	The registration provides administrative information according to ISO/IEC 11179. [--> Registered item]

2.4.2.7 Population instance

A population instance is a specific population provided in the context of a survey or other statistical activity instance. The population instance describes the population and its specific characteristics for a data collection provided within the statistical activity instance. The population instance describes an instantiation of the population, which is conceptually described in the related population definition.

One or more population instances can be referenced for defining the instantiation of the frame population (e.g. for persons, households and buildings).

Characteristic	Description
Identifier	A unique, language independent identifier is used to identify the population instance. This may typically be an abbreviation of its title or a systematic number. It is unique in the context of a population.
Title	The title is a short multilingual label for the population instance.
Description	This is a detailed multilingual description of the population instance, which typically describes the way the population instance is defined, the number of individuals and the time point or time period for which the population instance applies.
Population	The population that defines the population instance conceptually. [--> Population]
Data collection	When individuals of a population instance are collected in a data collection, the population instance can refer to this data collection. [--> Data collection]
Activity instance	The statistical activity instance, which creates the instantiation of the population, and in which the population instance is provided. [--> Statistical activity instance]
Registration	The registration provides administrative information in accordance with ISO/IEC 11179. [--> Registered item]

2.5 Related concepts

This section describe concepts, which are not directly relevant for defining variables, but which are linked to variables in one way or another. This includes foreign concepts, which are defined in other papers, administrative concepts, which are only touched on in this paper and some minor missing concepts, which have not been discussed in detail, but which are important, when implementing a metadata model based on the variable terminology model.

2.5.1 Foreign Concepts

Foreign concepts are concepts that have been defined in other places (e.g. *Neuchâtel group - Classifications*).

2.5.1.1 Classification family

Classification families are defined in more detail in the Neuchâtel Terminology Model Classification database object types and their attributes [1].

2.5.1.2 Classification

Classifications are defined in more detail in the Neuchâtel Terminology Model Classification database object types and their attributes [1].

2.5.1.3 Classification version

Classification versions are defined in more detail in the Neuchâtel Terminology Model Classification database object types and their attributes [1].

2.5.1.4 Classification level

Classification levels are defined in more detail in the Neuchâtel Terminology Model Classification database object types and their attributes [1].

2.5.1.5 Classification item

Classification items are defined in more detail in the Neuchâtel Terminology Model Classification database object types and their attributes [1].

2.5.1.6 Subject area

Subject areas are defined in more detail in the Reference Model [4].

2.5.1.7 Data collection

Data collection provides metadata for a particular collection of data items. Data collections are described by matrices (registers and cubes), which define the intension of the data collection, and by populations, which define the extension of the data collection. Moreover, the data collection refers to a number of datasets, which store the data produced within the statistical activity. All datasets related to the data collection are considered to contain the same data in terms of quantity and quality. Collection attributes describe the quality, methods and concepts for the data collection.

A data collection usually describes matrix data, i.e. a number of identically structured rows related to individuals or to aggregates of a given statistical unit type.

Data collections are typically produced as

- initial observation register
- final observation register
- cube
- table.

Data collections are defined in more detail in the Reference Model [4].

2.5.1.8 Data collection element

Data collection elements describe details resulting from the instantiation of contextual variables defined for the matrix, which describes the content of the data collection.

Considering a data collection as presenting a matrix with a number of columns, each data collection element describes the instance dependent metadata for a column in the matrix. The (more general) meaning of the column is described in the related contextual variable.

Such instantiation details described as data collection elements are quantities, completeness, bias or other (mainly quality information), which differ from data collection to data collection. Data collection elements may also define instance depending access conditions or regulations for sensitive data, when depending on the specific data collected.

Data collection elements are defined in more detail in the Reference Model [4].

2.5.1.9 Table instance

The table instance refers to the process generating the table and referring to the cube(s) the table is generated from. Table instances can be provided as generic table instances, which can refer to a number of tables derived from the same cube. All those tables are identified by one or more cube dimensions as year or community, which are fixed for each table generated from the generic table instance.

Table instances are defined in more detail in the Reference Model [4].

2.5.1.10 Record type

A record type describes a collection of record variables that are usually stored in a record of a statistical table, cube or register.

Record types are defined in more detail in the Reference Model [4].

2.5.1.11 Dataset

A dataset defines the type and physical location of the stored data for a record type. The output of the dataset can be pre-defined or the dataset information can be returned from the process. Since processes usually do not communicate with the metadata system, output datasets will be defined before running the process.

Data sets are defined in more detail in the Reference Model [4].

2.5.1.12 Footnote

A footnote may contain additional information about a classification, a classification version, a variable or many other metadata objects. Footnotes can be displayed in the context of tables and other publications.

Footnotes are defined in more detail in the Reference Model [4].

2.5.1.13 Keyword

Terms that are associated with one or more metadata objects are called keywords and can be used for keyword-based search. Metadata objects describing specific statistical concepts as classification or conceptual variable are described by means of topics. Keywords refer to specific topics (*e.g. classification or conceptual variable*) as well as to topics in general. Thus, it becomes possible to use those keywords for searching all topics associated with a keyword or associated topics of a specific topic type.

Keywords are defined in more detail in the Reference Model [4].

2.5.2 Administrative Concepts

Administrated or registered objects (items) have additional registration information, which is collected in a registered item. Registered items refer to maintenance units and suppliers and to user groups. They contain validation, creation and update information. Registered items are associated with a number of administrative object types described in this section.

Administrative object types describe entry points for internal and external users (e.g. *maintenance unit or subject areas*). The picture below shows administrative object types connected to a user group, which is responsible for maintaining the object type (link to user group), and object types registered by one or more persons that want to be kept informed about relevant changes in the related concept (link to person).

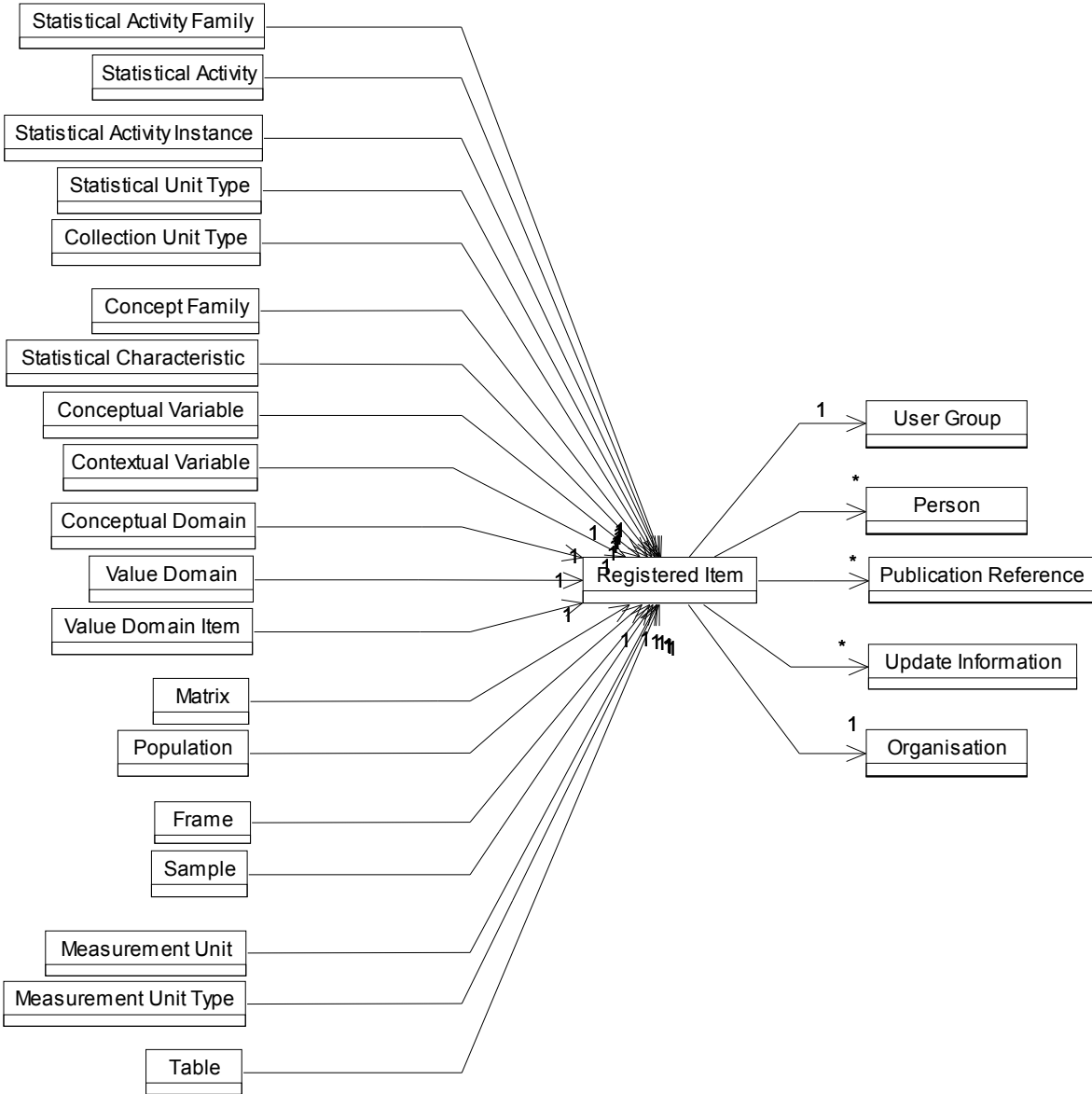


Figure 10 Registered item

2.5.2.1 Registered item

Registered items are registered in a metadata registry. In principle, any terminology object may behave as a registered item, but here, only the most important objects have been defined as registered items. Terminology objects, that can be registered, refer to the registered item via the registration characteristic.

Characteristic	Description
Created	Date on which the registered item was created.
Valid from	Date from which the registered item is valid.
Valid to	Date until which the registered item is valid.
Last update	Date and time of the last update in the registered item.
Standard	Marks the registered item as standard for preferred use.
Registration status	A designation of the position in the registration life cycle of an administered item, as described in ISO/IEC 11179-6. <i>Preferred standard, Standard, Qualified, Recorded, Candidate, Incomplete, Retired, Superseded.</i>
Administrative status	A designation of the position in the processing life cycle of a Registration Authority for handling registration requests. <i>Provisionally qualified, provisionally standard, provisionally preferred.</i>
Updates	A list of summary descriptions on updates that have occurred in the registered item. [--> Update information]
User Group	The group of persons within an organization that is responsible for the registered item, i.e. for updating or deleting it. [--> User group]
Registered users	List of registered users, who want to be informed when the registered item is updated or becomes invalid. [--> Person]
Organisation	The organisation responsible for maintaining the registered item. [--> Organisation]
Supplier	Organisation or unit that mainly delivers the registered item or can be asked to supply the registered item. When no supplier is registered, the registered item can be obtained from the maintenance organisation. [--> Organisation]
Sources	Documents or other sources, which provide a more detailed definition of the concept. [--> Publication reference]

2.5.2.2 Person

A person is an individual that acts within an organization or as a registered user for statistics. Each person is a contact person.

Characteristic	Description
Last name	Last name of the person.
First name	First name of the person. If a person has more than one first name all first names are displayed in this field.
Phone number	The person's current phone number.
Fax number	The person's current fax number.
Email address	The person's current email address.
Login name	The name the person uses to login to the registry
Password	Password the person has to use for login.
User groups	A person can be a member of any number of user groups. [--> User group]
Organisation	The organisation the person belongs to. [--> Organisation]

2.5.2.3 User group

The group of persons within an organization that is responsible for a registered item, i.e. for updating or deleting it.

Characteristic	Description
Identifier	Unique name/identifier of the user group.
Description	Description of the user group including its responsibilities.
Metadata objects	A list of registered metadata objects that the user group is responsible for. Objects referenced in this list may have different types. [--> Registered item]
Users	A list of users belonging to the user group. [--> Person]

2.5.2.4 Organisation

An organisation is a legal entity such as an institute, a company or a government institution. An organisation acts as an authority on many occasions.

Characteristic	Description
Identifier	The identifier is a unique short name for the organisation that is used to identify the organisation among others.

Characteristic	Description
Title	Official title or name the organisation is registered with or the way in which it is referred to in official documents.
Description	Short description of the organisation and its tasks.
Address	Postal address of the organisation (street number, city, zip code etc.)
Contacts	A number of contact people may be defined for the organisation. [--> Person]

2.5.2.5 Update information

Update information is associated with a certain metadata object in order to document an update.

Characteristic	Description
Timestamp	Date and time when the update has been made.
Description	Description of the background and reason for the update.

2.5.2.6 Publication reference

A publication is an official document or paper that has been published by an organisation.

Characteristic	Description
Title	The official title of the publication.
Type	The type of the publication describes the format in which the document was published <ul style="list-style-type: none"> - HTML for Internet browser - DOC for MS Word - PDF for Acrobat Reader document Typically, this is the extension used for software to access the document.
Location	Location refers to the place where the document is available. Depending on the document type this can be a file in the network, a WEB address or a publishing company.
Publication date	The publication date may consist of year, month and year, or the exact date the document was published.

3 Metadata object types by concept levels

This section describes metadata object types ordered by concept levels. Concept levels express metadata on different levels of abstraction. Concept levels provide different levels of abstraction, which allows defining metadata on different levels of detail. Concept levels not only differ in level of abstraction, but also in the scope, they are defined in. While metadata objects on the concept level provide common definitions in a statistical office, metadata defined on contextual level is valid in the context of a statistical activity (or survey). Instance metadata is defined in the context of a statistical activity or survey instance and metadata on the data level describes the way of accessing a specific data item or collection of data items.

While the **object variable** on the concept level defined the general concept of a variable in the office, the **contextual variable** provides a specialized definition in the context of a statistical activity. In principle, an object variable may relate to a number of contextual variables. In contrast to the **contextual variable**, the **data collection element** provides detailed information about the variable for a particular data collection, which has been produced within a specific activity instance. Metadata on the data level refers to particular data sets, which hold the data for a data collection.

The data level actually provides the link between metadata and data. Metadata on the instance level is not directly linked to data but indirectly via the data level. Metadata on the contextual level is linked indirectly to data via the instance level and metadata on the conceptual level via the contextual level.

Conceptual level

Concept family

Conceptual variable

Statistical characteristic

Statistical unit type

Fixed characteristic

Object variable

Conceptual domain

Measurement unit type

Measurement unit

Register unit type

Cube unit type

Aggregated unit type

Elementary unit type

Collection unit type

Classification

Classification version

Classification level

Classification item

Subject area

Contextual level

Statistical activity

Contextual variable

Register variable

Cube variable

Contextual value domain

Value domain item

Matrix

Register

Initial observation register

Final observation register

Cube

Table

Population

Target population

Frame population

Survey population

Instance level

Statistical activity instance

Frame

Sample

Data collection

Data collection element

Table instance

3.1 Conceptual level

The conceptual level is a metadata level, which describes statistical concepts of common interest. There are several metadata objects, which are of common interest and which are defined outside of any statistical activity.

The conceptual level provides a general-purpose definition for different statistical metadata objects. It provides also a mean for harmonizing statistical metadata using common concepts.

The conceptual metadata level does not relate directly to data but is linked to data via more specific metadata definitions on the contextual level.

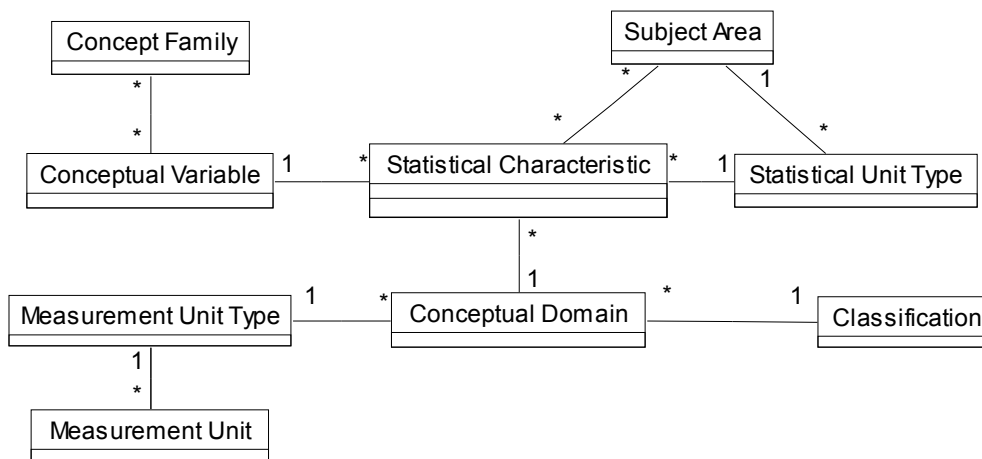


Figure 11 Conceptual level

Many metadata objects on the conceptual level do have appropriate metadata objects on the contextual level, where usually one conceptual metadata object relates to N contextual metadata objects. The table below shows the important relationships between conceptual and contextual levels:

Conceptual level	Contextual level
	Statistical activity
Concept family	
Conceptual variable	
Statistical characteristic	
Fixed characteristic	
Object variable	Contextual variable Cube variable Register variable
Statistical unit type	Matrix
Cube unit type	Cube
Register unit type	Register Initial observation register Final observation register
	Table
Conceptual value domain	Contextual value domain
Classification	
Classification item	Value domain item
Classification level	
Classification version	
Measurement unit type	
Measurement unit	
	Population Target population Survey population Frame population

3.2 Contextual level

The contextual level provides particular metadata definitions valid in a specific context, which is described as statistical activity. Contextual metadata data usually refers to general metadata concepts, but contains more specific (context related) definitions (*e.g. for variables or value domains*).

On the contextual level, ideas and concepts about statistical products can be described more formalized, but also without any direct connection to data. Contextual metadata is created before producing statistics to describe the statistical product to be achieved. Thus, contextual metadata is also not directly related to data, but to metadata on the instance level.

The context, in which contextual metadata is defined, is described by a statistical activity or statistical survey. Within a statistical activity, metadata becomes more specific than metadata being defined on the concept level. Moreover, a statistical activity forms a proprietary namespace, which allows naming metadata objects within the activity without conflicting with names in other statistical activities.

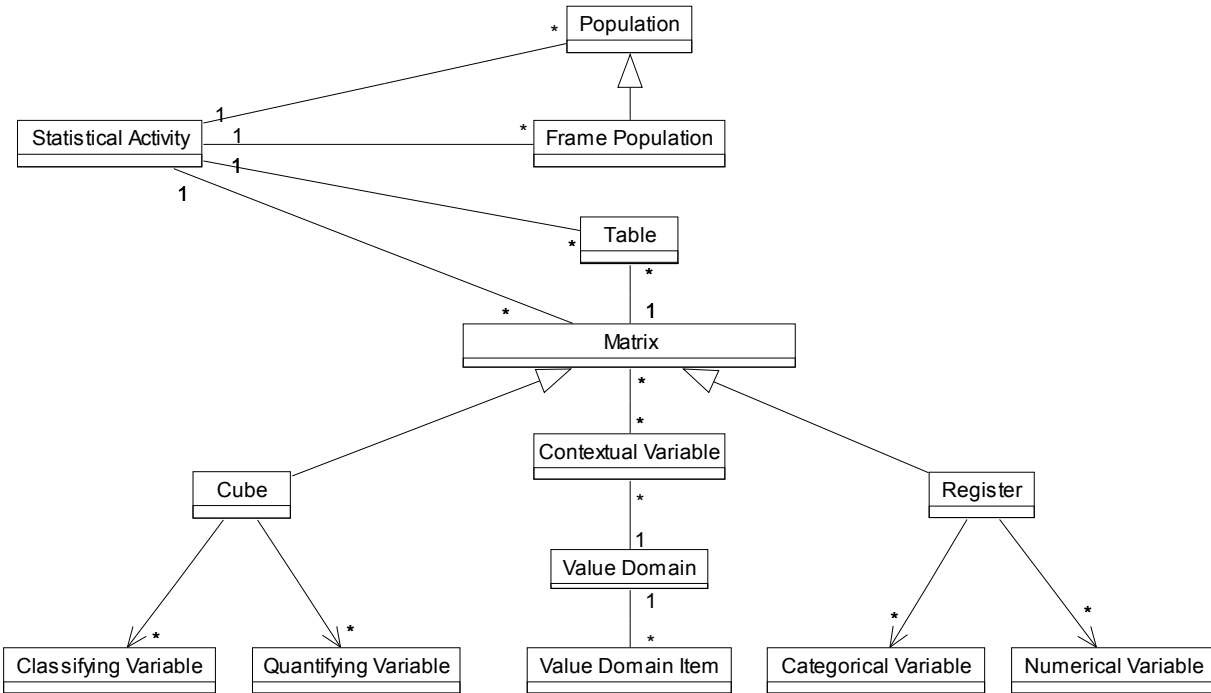


Figure 12 Contextual level

Metadata objects on the contextual level do have appropriate metadata objects on the instance level, where usually one contextual metadata object relates to N instance metadata objects. The table below shows the important relationships between contextual and instance level:

Contextual level	Instance level
Statistical activity	Statistical activity instance
Contextual variable Cube variable Register variable	Data collection element
Contextual value domain	
Value domain item	
Matrix Cube Register Initial observation register Final observation register	Data collection
Table	Table instance
Population Target population Survey population Frame population	Sample Frame

3.3 Instance level

The instance level describes instantiations of contextual metadata objects. When running a survey or creating a statistical product, the concept described in the statistical activity, is instantiated within an activity instance (survey instance).

Subject to definition on the instance level are particular statistical products and associated data sets or tables. Usually there is a 1:N relationship between metadata objects on the contextual and on the instance level (*e.g. statistical activity: activity instance*).

On the instance level, specific information about running the survey or product is stored, quality metadata is provided and data collections and table instances are described. The instance level is tightly linked to the data level, which provides the direct data link.

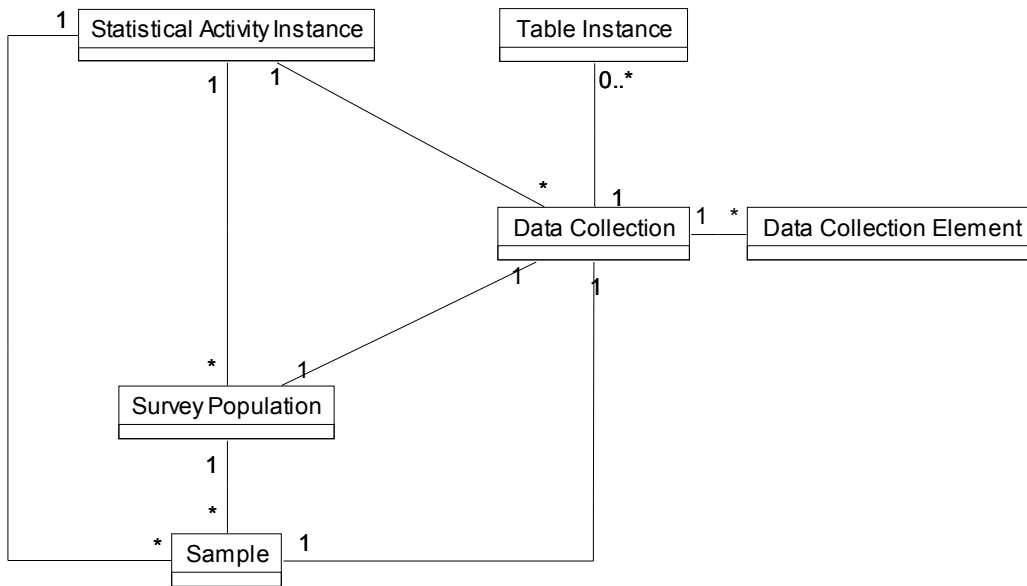


Figure 13 Instance level

The only link between the instance and the data level is the link between data collections and datasets. This is sufficient to provide extensive metadata for each single data item.

4 References

- [1] Neuchâtel Terminology Model Part II (Classifications and related objects)
August 2004, Version 2.1
http://ec.europa.eu/comm/eurostat/ramon/miscellaneous/index.cfm?TargetUrl=DSP_STANDARD
- [2] ISO/IEC 11179: 2005 Metadata registries (All parts). Geneva: International Organization for Standardization and International Electrotechnical Commission.
- [3] ISO/FDIS 1087-1: 2000 Terminology - Part 1: Vocabulary. Geneva: International Organization for Standardization.
- [4] Reference Model (TM)
September 2005, Version 3
www.run-software.com/ReferenceModel
- [5] A terminology model approach for defining and managing statistical metadata
Reinhard Karge, April 2005, Open Forum Berlin
<http://www.berlinopenforum.de/download.html>

5 Annexes

5.1 Annex I - Notation and multiplicity in figures

The following table gives an overview of the notation used in our figures.

Notation	Explanation	Examples
A — B	Association between A and B.	<i>An employee is employed in an establishment. Association: employment</i>
A —> B	Association from A to B. Unidirectional navigation.	<i>Obtain contact information for an employee.</i>
A —▷ B	A inherits from B. B is a generalisation of A. A is a specialisation of B.	<i>A shape is a generalisation of a circle. A circle is a specialisation of a shape.</i>
A —◊ B	Aggregation. B is an aggregation of A.	<i>An enterprise is an aggregation of establishments</i>

The next table gives an overview of the multiplicity notation used in our later figures.

Multiplicity	Explanation
1	Null or one.
*	Null, one or more

5.2 Annex II - Defining terminology models

Terminology models are defined by specifying concepts for constructs, *e.g. statistical variable, statistical classification*, and its attributes and relationships to other concepts. Such a two level terminology definition corresponds directly to a conceptual metadata model where the concepts are considered as metadata object types and the attributes and relationships of these concepts are considered as characteristics of those metadata object types.

The meta-model for a terminology helps to bridge the gap between subject area experts and IT experts . The meta-model tries to refer as much as possible to available standards. However, to provide a structured presentation of knowledge, it needs further extensions.

The following picture shows the complete set of terminology objects and their relationships. Details of the meta-model are described in [3].

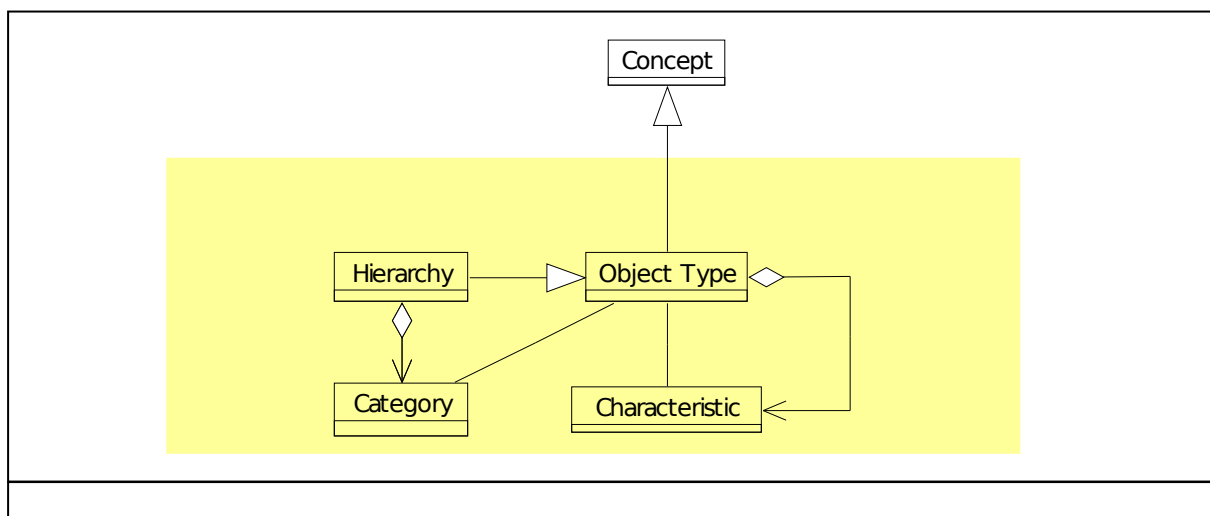


Figure Concept extensions for terminology model

The Neuchâtel Terminology model is an instance of the terminology meta-model. The model contains object types and characteristics for statistical variables and statistical classifications. This is an essential part of conceptual definitions and provides a view to statistical knowledge structures.

The terminology meta-model mainly describes the *object type - characteristic – object type* relations. A relevant concept of an expert area is considered as an object type, which has characteristics. Characteristics refer to object types again, which then must be described by characteristics etc.

Even though this is a simplified definition of the world, the terminology provides a good conceptual definition of many relevant concepts on the one hand, and a basis for generating database models and other technical information on the other hand.

A small extension of the static terminology model includes hierarchies, which are often necessary for defining categories of object types defined in the model.

The characteristics in the model should be described according to its categories, since it makes a difference whether a company *is* a person or *has* a person. If a company were a person, the company would inherit all the characteristics that are defined for person.

Defining a terminology model means defining the concepts for a subject field. The Neuchâtel Terminology Model refers to the subject field of statistics. The specific subjects of statistical variables and statistical classifications are defined.

5.2.1 Concept

A concept is "a unit of knowledge created by a unique combination of characteristics" (ISO/FDIS 1087-1). Concepts have a set of characteristics, the intension, which are used to describe the different object types within the terminology model. Concepts are designated by terms and may have synonyms. Terms (including synonyms) for concepts must be unique in the subject area the concept belongs to or within an object type.

Synonyms: List of synonyms that can be used instead of the concept name.

Example: One or more examples describing the concept. Examples are marked by italic.

Characteristic	Description
Name	The name is a single word or group of words that designates the concept. Synonym: term
Definition	Representation of a concept by a descriptive statement that serves to differentiate it from related concepts [ref. 3]

5.2.2 Object type

The object type is a specialization of general concept as defined in ISO/FDIS 1087-1. An object type is either complex (one or more characteristics specified) or elementary (no characteristics specified). Text and number are typical elementary object types. An object type can also be a hierarchy.

Characteristic	Description
Concept	The object type inherits all the characteristics of the concept. [--> Concept]
Characteristics	List of characteristics that describe details of an object type. [--> Characteristic]
Categories	Categories associated with the object type. [--> Category]

5.2.3 Characteristic

A characteristic or concept relation belongs to an object type. The designations for characteristics must be unique within the scope of the object type. Characteristics sometimes are represented as object types themselves. In this case, we say that characteristics have characteristics.

Synonym: Attribute

Characteristic	Description
Name	The name is a single word or group of words that designates the characteristic.
Definition	Representation of a characteristic by a descriptive statement that serves to differentiate it from related characteristics [ref. 3].

Characteristic	Description
Object type	Object type to which the characteristic belongs and that defines the characteristics or categories for the characteristic. [--> Object Type]

5.2.4 Hierarchy

The hierarchy provides a schema for partitioning a given class of objects into subclasses. A hierarchy contains a number of categories, and it may also contain characteristics the categories consist of. Defining hierarchy becomes necessary when, conceptually, objects must be partitioned by categories.

Characteristic	Description
Object type	Conceptually, a hierarchy is described as an object type and inherits all the characteristics of the object type. [--> Object Type]

Hierarchies are not explicitly defined in the variable terminology model, but they are defined implicitly in several places (*e.g. statistical unit type/component types*).

5.2.5 Category

A category is a component of a hierarchy, and it is used as a subclass of a given class. The set of categories partitions the given class. For instance, the object defined by *all persons in Norway in 2006* may be classified by *sex*. Therefore, *sex* is a hierarchy. Also, *male* and *female* are categories of that hierarchy that enable the breakdown of the population of objects into subsets of male persons and female persons.

Characteristic	Description
Name	The name is a single word or group of words that designates the category.
Definition	Definition or description of the category.
Sub-divisions	One or more subordinated hierarchies can be provided, which define subsequent division of the class into subclasses, i.e. objects of one category can be divided in subclasses by means of different hierarchies. [--> Hierarchy]
Object type	Object type, which describes the intensional concept of the category, i.e. the intensional definition of objects in the class defined by the category. [--> Object type]

5.3 Annex III - Comparison with ISO/IEC 11179 Metamodel

This annex contains a comparison of part of the Terminology Model with the metamodel specified in Part 3 (2nd edition - 2003) of ISO/IEC 11179 (hereafter referred to as 11179-MM). The comparison is limited to that part of the Terminology Model described in section 2.3 (Variable Structure). We refer to this as TM-2.3.

The 11179-MM has 4 basic classes, as illustrated in Figure A-1 below:

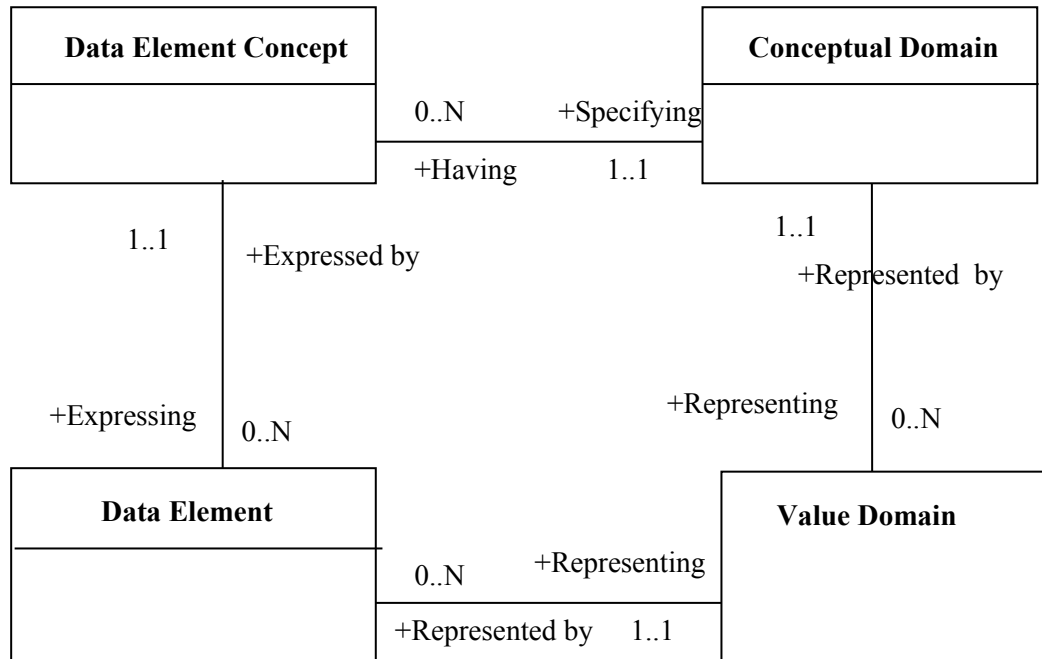


Figure 11179-MM Overview

Each class in the 11179-MM is defined as follows in the standard:

- **Data Element**: unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes
- **Data Element Concept**: concept that can be represented in the form of a data element, described independently of any particular representation
- **Conceptual Domain**: set of valid value meanings
- **Value Domain**: set of permissible values

A permissible value is an ordered pair consisting of a value and its meaning - the value meaning. This way, each value in a Value Domain always has one meaning associated with it.

A Data Element Concept is made up of two other concepts:

- **Object Class**: set of ideas, abstractions, or things in the real world that are identified with explicit boundaries and meaning and whose characteristics and behaviour follow the same rules
- **Property**: characteristic common to all members of an object class

Comparison

The following is a mapping between the TM-2.3 and the 11179-MM. In the table below, there are two columns, one for the TM-2.3 and the other for 11179-MM. Classes listed together in the same row are similar, but not exactly the same. Classes that are listed alone in either column have no direct equivalent.

TM-2.3	11179-MM
Statistical unit type	Object Class
Conceptual variable	Property
Statistical characteristic	
Fixed characteristic	
Object variable	Data Element Concept
Conceptual domain	Conceptual Domain
Value domain	Value Domain
Contextual variable	Data Element
Concept family	Concept ⁸
	Representation Class
Classification item ⁹	Value Meaning

Statistical Characteristic does not have an equivalent in 11179-MM because of the Fixed Characteristic subtype. There is no equivalent of a Fixed Characteristic in 11179-MM.

11179-MM has the Representation Class, which has no equivalent in TM-2.3. This class is defined as follows: hierarchy of types of representations. For statistical data, this hierarchy is the hierarchy of statistical data types as defined in the following example:

- Text
- Categorical
 - Ordinal
 - Nominal
- Quantitative (level A and B provide possible sub-categories)
 - Level A
 - Discrete
 - Continuous
 - Level B
 - Interval
 - Ratio

11179-MM has the Value Meaning class, which also has no equivalent in TM-2.3. It corresponds, however, to the classification items of a classification version the conceptual domain is linked to. Classifications and related objects are not the main subject of the variable model but are defined in the Neuchâtel Classification model.

Finding values that mean the same as other values is a necessary for data harmonization. Managing the Value Meaning allows one to find all equivalent values easily. An example of where value equivalence is important is multiple codings of a single set of concepts. The standard ISO 3166-1 *Country Codes* contains three different codings for all the identified countries in the world: 3 character

⁸ The notion of a general Concept class is being added to ISO/IEC 11179 and will appear in the 3rd edition of the standard.

⁹ Value meanings can be defined by means of items of a classification by linking the conceptual domain to a classification version or a specific level in a classification version. This, however, does not necessarily provide value meanings for any value domain.

alpha codes, 2 character alpha codes, and 3 number numeric codes. So, the codes for Afghanistan, the first country in the list, are as follows: AfG, AF, and 004.

There are minor distinctions between a Data Element Concept and an Object Variable. The concept of a variable in statistics is very much associated with the characteristic of a population. The characteristic is what analysts pay close attention to. The specific population may not be explicit. TM-2.3 reflects this view.

Summary

11179-MM and TM-2.3 touch similar statistical metadata areas, but from different perspectives. ISO/IEC 11179 is a general description of data, independent of subject matter. It also supports registration, a methodology for administering content including its quality. TM-2.3, on the other hand, describes and classifies statistical data. Because its context is statistics, TM-2.3 contains concepts and terminology familiar to statisticians.

Interestingly, the structure of both models is remarkably similar, which the comparison chart above shows. This similarity allows exchanging metadata between ISO/IEC 11179 registries and TM-2.3 repositories¹⁰. Thus, one may generate an ISO/IEC 11179 metadata registry from a TM-2.3 compliant metadata system, and vice-versa.

Combining the administrative components of ISO/IEC 11179 with the detailed, subject matter specific, content definitions of TM-2.3 provide the basis for a standards-conformant¹¹ statistical metadata registry. This covers both registration aspects and rich content definitions for statistical offices.

¹⁰ Metadata registries and metadata repositories are both databases of metadata. However, a metadata registry also supports the functionality of registration.

¹¹ Conformance is a specific term in standards work. It means, roughly, any implementation of a standard is conformant if it satisfies all the requirements in the standard.

5.4 Annex IV - Examples of concepts listed alphabetically by concept

Concept	Example
Aggregated unit type	<i>Trade sector</i>
Classification	<i>Nomenclature generale des Activites economique dans les Communautés Europeenes (NACE)</i>
Classification item	<i>50 Sale, maintenance and repair of motor vehicles and motorcycles, retail sale of automotive fuel</i>
Classification level	<i>Level 3: Division</i>
Classification version	<i>NACE revision 3</i>
Collection unit type	<i>Establishment</i>
Concept family	<i>Economic performance of businesses</i>
Conceptual domain	<i>Standard for industrial classifications (SIC)</i>
Conceptual variable	<i>Turnover, employment</i>
Value domain	<p><i>SIC-categories:</i></p> <p><i>50 Sale, maintenance and repair of motor vehicles and motorcycles, retail sale of automotive fuel</i></p> <p><i>51 Wholesale trade and commission trade, except of motor vehicles and motorcycles</i></p> <p><i>52 Retail trade, except of motor vehicles and motorcycles; Repair of personal and household goods</i></p>
Contextual variable	<i>Kind of activity of establishments in the wholesale and retail trade</i>
Cube	<i>Principal figures (employment, turnover, compensation of employees, investments) by industry subclass (division) for establishments is one of the cubes, resulting from the statistical activity structural statistics, wholesale and retail trade of Statistics Norway.</i>
Cube unit type	<i>A cube unit type can be an elementary or aggregated unit type. The statistical activity structural statistics of wholesale and retail trade of Statistics Norway generates cubes, in which the trade establishment is the elementary unit type, and the trade sector is the aggregated unit type.</i>
Cube variable	<i>Kind of activity, number of employees, number of establishments, compensation of employees, turnover.</i>
Data collection	<i>Final Observation Register for the Structural Surveys of Statistics Norway for the financial year</i>

Data collection element	<i>Quality information; e.g. imbalances in the sample (e.g. misleading stratification) can cause errors for the variables for which information has not been obtained for all units in the population</i>
Elementary unit type	<i>Establishment</i>
Final observation register	<i>Final Observation Register for the Structural Surveys of Statistics Norway for the financial year</i>
Fixed characteristic	<i>51.874 is the SIC- category that, applied to the object variable kind of activity of the statistical unit type establishment results in defining the subtype SIC 51.874 establishment</i>
Frame population instance	<i>Establishments in the wholesale and retail trade in 2002(, including relevant information about their addresses etc.) belonging to The Norwegian Central Register of Enterprises and Establishments</i>
Initial observation register	<i>Initial Observation Register for the Structural Surveys of Statistics Norway for the financial year</i>
Matrix	<i>Final Observation Register for the Structural Surveys of Statistics Norway for the financial year</i>
Measurement unit	<i>Norwegian Crown</i>
Measurement unit type	<i>Currency</i>
Object variable	<i>Applying the conceptual variable of employment to the statistical characteristic having employees of the statistical unit type establishment generates the object variable number of employees of establishments.</i>
Population Instance	<i>All establishments existing at any moment in year 2002, and engaged in wholesale or retail trade as their main kind of activity</i>
Frame	<i>The Norwegian Central Register of Enterprises and Establishments is the frame for the statistical activity structural wholesale and retail trade statistics of Statistics Norway</i>
Frame instance	<i>The Norwegian Central Register of Enterprises and Establishments in 2002</i>
Register	<i>Final Observation Register for the Structural Surveys of Statistics Norway for the financial year</i>
Register unit type	<i>Establishment</i>
Register variable	<i>Number of employees, kind of activity, turnover.</i>
Statistical activity	<i>Structural wholesale and retail trade statistics of Statistics Norway</i>
Statistical activity family	<i>Structural statistics of Statistics Norway</i>
Statistical activity instance	<i>Structural statistics for the wholesale and retail trade of Statistics Norway in 2002</i>
Statistical characteristic	<i>Turnover of the statistical unit type establishment</i>
Statistical unit type	<i>Establishment</i>

Subject area	<i>Industrial activities</i>
Table instance	<i>Principal figures by size class (number of employees) and industry subclass (division) for 2002.</i>
Value domain item	<i>50 Sale, maintenance and repair of motor vehicles and motorcycles, retail sale of automotive fuel</i>