



UNITED NATIONS
DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS
STATISTICS DIVISION

ESA/STAT/AC.289/22
11 May 2015

**Meeting of the Expert Group on
International Statistical Classifications
New York, 19-22 May 2015**

Generic Statistical Information Model (GSIM): Statistical Classifications Model

Reprint of UNECE document

Generic Statistical Information Model (GSIM): Statistical Classifications Model

(Version 1.1, December 2013)

About this document

This document defines the key concepts that are relevant to structuring statistical classification metadata, and provides the conceptual framework for the development of a statistical classification management system. It is aimed at classification experts.



This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>. If you re-use all or part of this work, please attribute it to the United Nations Economic Commission for Europe (UNECE), on behalf of the international statistical community.

Contents

Summary	3
1 Introduction	3
1.1 Background.....	3
1.2 Context and scope	4
1.3 Classification and related concepts	5
1.4 Other terminologies	8
1.5 Implementation	9
1.6 Layout of the GSIM Statistical Classification Model.....	10
1.7 References	10
2 Object types	11
3 Object graph	12
4 Object types and attributes	13
4.1 Classification Family	13
4.2 Classification Series.....	13
4.3 Statistical Classification	14
4.4 Level.....	17
4.5 Correspondence Table.....	17
4.6 Classification Index	18
4.7 Classification Item	19
4.8 Map	21
4.9 Classification index entry.....	22
Appendix 1: Worked example of the GSIM Statistical Classification Model	23
Appendix 2: Checklist of possible content for the Introduction to a Statistical Classification	36
Appendix 3: A typology of Classification Item changes	39

Summary

1. The GSIM Statistical Classifications Model is based upon the Neuchâtel terminology model for classification database object types and their attributes v2.1. It was developed by a group of 19 members from 13 different national and international organisations in an endeavour to arrive at a common language and a common perception of the structure of statistical classifications and the links between them. The GSIM Statistical Classifications Model is both a terminology and a conceptual model. It defines the key concepts that are relevant to structuring Statistical Classification metadata and provides the conceptual framework for the development of a Statistical Classification management system.
2. The model has a two level structure, consisting at the first level of the object types (e.g. Statistical Classification, Classification Item), 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 object types and attributes which are related solely to the technical aspects of Statistical Classification management. The conceptual model is generally applicable and not dependent on IT software and platforms. It may be used in any context where structured information on Statistical Classifications is required. A simplified object graph gives an overview of the main object types and relationships in the conceptual model.

1 Introduction

1.1 Background

3. In June 1999, a meeting on terminology was held in Neuchâtel, Switzerland, with participants from the statistical offices of Denmark, Norway, Sweden and Switzerland and the software developers in Run Software-Werkstatt. This was the start of the "Neuchâtel group". The aim of the group was to clarify some basic concepts and to arrive at a common terminology for classifications. The terminology defined the key concepts that were relevant for how to structure classification metadata and provided the conceptual framework for the development of a classification database. The work listed and described the typical object types of a classification database, and the attributes connected with each object type.

4. The development of the model had a practical focus as all of the participating National Statistical Organisations (NSOs) planned to use it in their own implementation of a classification database. The most important purposes for developing a classification database were:

1. to make accessibility and maintenance of classifications easier, and
2. to ensure common use of classifications across different fields of statistics.

A central database was the preferred solution because it realised one of the important principles of metadata - document and update once (centrally), and reuse wherever it is relevant. The Neuchâtel terminology model: Classification database object types and their attributes (version 2.0) was released in 2002.

5. Later, Statistics Netherlands joined the Neuchâtel group, and a new version of the terminology, version 2.1 (with one new object and one new attribute), was released in 2004.

6. It was essential for the Neuchâtel group that the terminology should be flexible and independent of IT software and platforms. This resulted in different classification database implementations for the participating NSOs, according to specific needs and policies. Also, it was always an important premise for the group that the work should be public and available to anyone free of charge.

7. Many countries have at least partially implemented the model¹. After years of practical experience, several of the implementing countries expressed a desire to see some revisions to the model. As the Neuchâtel group no longer existed, a possible revision was discussed at the 2011 METIS Workshop². Subsequent to the workshop, the METIS Steering Group contacted the UN Expert Group on International Statistical Classifications to work on the revision of the Neuchâtel model. As a result, a joint working group was created, bringing together classification and statistical metadata experts.

8. At the same time, a project sponsored by the High Level Group for the Modernization of Statistical Production and Services was reviewing the Generic Statistical Information Model (GSIM)³. GSIM provides the information object framework supporting all statistical production processes such as those described in the Generic Statistical Business Process Model (GSBPM)⁴, giving the information objects agreed names, defining them, specifying their essential properties, and indicating their relationships with other information objects. In the development of GSIM, the objects related to classifications were mostly drawn from the Neuchâtel Terminology Model.

9. During the revision work it was discussed and decided that for the future the Neuchâtel model for classifications will be part of GSIM. Several objects and attributes have been changed during the revision process, and the revised model will in practice be an annex to GSIM.

1.2 Context and scope

10. Classifications are generally regarded as a special kind of metadata for statistics. They are definitional, content-oriented metadata, ordering and describing the meaning of statistical data. A classification database can be described as a register of meta-objects (classifications and related object types), which in turn have their own set of metadata. It forms a more or less technically integrated part of the overall metadata information system of a statistical office.

11. The GSIM Statistical Classifications Model orders the concepts in a two-level structure of object types and attributes. On the first level, it specifies the basic object types of a classification database (e.g. Classification Family, Classification Series, Statistical Classification, Correspondence Table, Classification Index) and, on the second level, it lists the attributes connected with each object type. It is both a terminology and a conceptual model. It provides the conceptual framework for the development of a classification database. This immediate practical purpose has obviously limited its scope. It is not concerned with

¹ Countries that have implemented the model include Austria, Belgium, Bulgaria, Canada, Croatia, Czech Republic, Denmark, Estonia, Germany, Greece, Ireland, Norway, Portugal, Slovak Republic, Slovenia, Sweden, Switzerland and the Netherlands

² “METIS” was the joint UNECE / Eurostat / OECD group on statistical metadata

³ See: <http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model>

⁴ See: <http://www.unece.org/stats/gsbpm>

recording all the terms used in this area, nor does it deal with methods or best practices in the development and management of classifications. What it does do, is define the key concepts that are relevant for how to structure classification metadata and, indirectly, how to present information on classifications to different kinds of users. Since the GSIM Statistical Classifications Model belongs to the semantic and conceptual sphere of metadata it does not include object types and attributes that are related solely to the technical aspects of a classification database.

1.3 Classification and related concepts

12. According to ISO 704: 1987 (E) Principles and methods of terminology, a term is a word or phrase, which designates a concept. This section sets out some central concepts related to classifications and the relationships between them and the terms that will be used to refer to these concepts.

13. In the field of statistics, the term **classification** is normally used to denote one of the following concepts:

- a) The general idea of assigning statistical units to categories representing the values of a certain variable.
- b) The general concept of a structured list of mutually exclusive categories, each of which describes a possible value of the classification variable. Such a structured list may be linear or hierarchically structured. A linear classification is a list of categories, which are all at one and the same level (e.g. the ISO 3166 country code list, or a classification of marital status). In a hierarchical classification the categories are arranged in a tree-structure with two or more levels, where each level contains a set of mutually exclusive categories. The items of each level but the highest (most aggregated) are aggregated to the nearest higher level. In common usage the term classification often implies a hierarchical classification.
- c) One particular structured list of mutually exclusive categories, which is named, has a certain stability and normative status, and is valid for a given period of time (e.g. ISIC Rev.1).
- d) One particular named set of several structured lists of mutually exclusive categories, which are consecutive over time and describe the possible values of the same variable (e.g. ISIC).

14. The distinction between concepts c. and d. is seldom made explicit. Here as well, the term rather implies a hierarchical classification, and especially one of the group of "large", traditional, well-established and normative standard classifications.

15. **Nomenclature** is a term, which is closely related to classification. Nomenclature has to do with naming. Basically it denotes a list of named entities. Adding system and structure to the list turns it into something that resembles a classification. Although they do not have exactly the same meaning, the terms classification and nomenclature are often regarded as synonyms and used interchangeably. Nomenclature is not a term used in GSIM terminology.

1.3.1 Classification Series and Statistical Classifications

16. The conceptual framework of the classification database includes an object type roughly equivalent to concept d above. In the GSIM Statistical Classifications Model, this

concept has been named **Classification Series**. The concept of each "structured list of mutually exclusive categories" has been named **Statistical Classification**.⁵

17. A Statistical Classification is a set of categories (Classification Items) which may be assigned to one or more variables registered in statistical surveys or administrative files, and used in the production and dissemination of statistics. The Classification Items are defined with reference to one or more characteristics of a particular population of units of observation. A Statistical Classification may have a flat, linear structure or may be hierarchically structured, such that all Classification Items at lower Levels are sub-categories of a Classification Item at the next Level up. The Classification Items at each Level of the classification structure must be mutually exclusive and jointly exhaustive of all objects in the population of interest.

1.3.2 Some attributes related to statistical classifications

18. Statistical Classifications vary in their relationship to other Statistical Classifications. The following paragraphs discuss the terms that relate to such variation.

Classification version

19. A Statistical Classification is a version if it has a certain normative status and is valid from a particular date for a period that may or may not be specified. A new version is created when such a Statistical Classification is superseded by the introduction of a new Statistical Classification that differs in essential ways from the previous version. Essential changes are changes that alter the borders between categories, i.e. a statistical object/unit may belong to different categories in the new and old versions. Border changes may be caused by creating or deleting categories, or moving part of a category to another. These changes can occur at any Level of the classification. The addition of case law, changes in explanatory notes or in the names of Classification Items do not lead to a new version.

20. It should be noted that if a Statistical Classification is superseded by a new version, the two versions will likely serve the same objective or purpose.

21. Statistical Classifications that are related to each other as versions belong to the same Classification Series.

Classification variant

22. A particular Statistical Classification may not meet all the needs of its users. If it is for dissemination or other uses, the classification structure may be ill suited for the purpose at hand (for example, Levels or categories are too general or too narrow, too detailed in one area, and too broad in another). To meet these needs, a number of alternatives may be created, in which the original categories are split or regrouped to provide context-specific additions or alternatives to the standard aggregation structure. These are called classification variants.

⁵ Classification Series corresponds to Classification in the Neuchâtel terminology model. Statistical Classification includes Classification Version and Classification Variant from the Neuchâtel model.

23. A Statistical Classification is a variant of another Statistical Classification if it is built from the Classification Items of that base Statistical Classification. These Classification Items do not need to be all at the same Level in the base statistical classification. To these Classification Items, one or more new Levels may be added. This can include extending the base Statistical Classification with one or several new Levels at the bottom of its base, creating a new lowest Level. It should be noted that variants are typically developed to serve a specific purpose.

24. Variants are commonly of three kinds. These have been named **extension variants**, **aggregate variants** or **regrouping variants**. There could exist other types of variants. A particular variant could include elements from more than one of these variant types.

25. **Extension variant:** An extension variant is a Statistical Classification that extends the base Statistical Classification with one or several new Levels at the bottom, creating a new lowest Level. An extension variant thus adds new lower Levels to the base Statistical Classification but does not otherwise alter its original structure.

26. **Aggregate variant:** An aggregate variant is a Statistical Classification that groups the categories of a linear Statistical Classification to create one or several aggregate level(s), thus creating a hierarchy.

27. **Regrouping variant:** A regrouping variant is a Statistical Classification that introduces additional or alternative aggregate levels by regrouping categories of the base statistical classification. Two types of regrouping variants have been identified:

a) Regrouping variants which do not violate the structure of the base Statistical Classification: This type of regrouping variant introduces a new level or new levels on top of, or in between existing Levels of a hierarchical Statistical Classification without otherwise altering the original structure of the hierarchy. This regrouping variant consists of all classification Levels of the base Statistical Classification plus the new variant Level(s). The parent Level (if any) of the new variant Level can be either another variant Level or a Level from the base Statistical Classification.

b) Regrouping variants which violate the structure of the base Statistical Classification: This type of regrouping variant introduces a new Level or new Levels on top of any but the topmost Level of a hierarchical Statistical Classification by regrouping categories of the base Statistical Classification in a way which violates its original order and structure. This regrouping variant consists of all classification Levels of the base Statistical Classification below the new variant Level(s) plus the new variant Level(s). In such a regrouping variant, a new variant Level cannot have a base Statistical Classification Level as parent Level.

28. In all variants except regrouping variants which violate the structure of the base Statistical Classification, all Levels of the base Statistical Classification are retained and one or more new Levels are inserted. In regrouping variants which violate the structure of the base Statistical Classification, one or more new Levels are inserted and only the base Statistical Classification Levels below the new variant Levels are retained.

29. It is sometimes debated whether a classification database should be descriptive or prescriptive, the idea being that a prescriptive database will contain only standard

classifications, whereas a descriptive database will also contain non-standard variants. In reality, the demarcation between standard and non-standard classifications or between these and more loosely structured groupings is not very clear. It seems, therefore, that the criterion for inclusion in the database cannot be formal status only, but just as much the usefulness and commonality of the information provided. Most of the time the departures from the norm are legitimate, made to meet specific producer requirements or user needs. In any case alternative groupings exist and have to be documented. Indeed, listing the non-standard variants used in a statistical office may be a first and necessary step towards reducing their numbers.

Classification update

30. A Statistical Classification is an update of another Statistical Classification if it supersedes that Statistical Classification but does not differ from it in essential ways. Updates to specific elements of a Statistical Classification may be permissible within the life of a version. They may simply be noted in the context of the element affected or, if the changes are sufficiently numerous or significant, a new Statistical Classification can be issued that supersedes the previous Statistical Classification.

Floating classification

31. A Statistical Classification is said to be floating if it permits updates and essential changes without requiring their recognition through the issuing of a new Statistical Classification. Such Statistical Classifications may be used, for example, in contexts where change in the variable is expected to occur, but irregularly, and such change must be incorporated into the Statistical Classification in a timely fashion. Dates of validity on all elements of these Statistical Classifications allow the reconstruction of the Statistical Classification as it was on any particular date.

1.4 Other terminologies

32. There exist a number of terminologies and glossaries dealing with classification terms. These are either concerned with metadata in general or more specifically focused on classifications. The UN Glossary of Classification Terms is a multi-purpose general glossary of concepts, which also contains information on actual classifications and best practices in the development of classifications. It is much broader in scope than the GSIM terminology.

33. The draft Glossary of Statistical Terms attached to the joint OECD and Eurostat SDMX paper Developing a Common Understanding of Standard Metadata Components draws heavily on the UN glossary for its classification related terms.

34. There is also the UNECE "METIS" Terminology on Statistical Metadata. This has the term classification scheme instead of Classification Series but the concept is the same.

35. Concepts and terms related to classifications are also found in more general papers, for example, Best Practice Guidelines for Developing International Statistical Classifications, a paper developed by the UN Expert Group on International Statistical Classifications. This paper describes best practices for the development, use, maintenance and revision of classifications and there is close alignment with the GSIM Statistical Classification Model. The usage and scope of the best practice document are, however, different from those of the GSIM Statistical Classification Model.

36. The GSIM Statistical Classification Model terminology should be regarded as a complement rather than a rival to other terminologies in the field. Naturally there is a certain overlap of terms with the glossaries and papers mentioned above. In most cases there is a general agreement between the concepts and the terms used, although the wording of the definitions may vary. Not surprisingly, the one instance in which the terminology is at variance with other terminologies is in using the term "Statistical Classification" for one particular and well defined concept, and for making a clear distinction between "Classification Series" and "Statistical Classification" as explained in previous paragraphs. This and a few other instances of inconsistency are due to the particular focus and purpose of the GSIM terminology, which calls for quite specific and narrowly, defined concepts.

1.5 Implementation

37. Although the original Neuchâtel terminology was initially developed in the context of the classification database application of Bridge^{NA}, 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 classifications is needed.

38. In the context of the Bridge^{NA} system the conceptual framework has been used to develop a general semantic interface for metadata (ComeIn). It has also served as a specification for a concept-guided and user-oriented dialogue application, which functions as a browser and a tool for the management of classifications. This application is used in Statistics Sweden with the aim of setting up, developing and managing their national classification database.

39. In 2013, a questionnaire investigated the use of standards relevant to classifications and the need for a revised Neuchâtel Model for Classifications. Responses were received from eighteen countries or international organisations: Australia, Austria, Canada, Croatia, Estonia, France, Germany, Ireland, the Netherlands, New Zealand, Norway, Portugal, Slovenia, Sweden, Switzerland, the United States, Eurostat and the ILO.

40. Table 1 contains a summary of the results regarding the use of the Neuchâtel Model for Classifications.

Table 1: Summary of results

Standards and models	Considering Use	Currently in use
Neuchâtel Model for Classifications	4	11
<i>Neuchâtel terminology</i>		
Classification family	3	13
Classification	2	14
Classification version	2	14
Classification variant	2	9
Classification index	3	10
Correspondence table	2	14
Classification level	2	14

Classification item	2	14
Item change	1	7
Case law	3	4
Classification index entry	3	10
Correspondence item	2	14

1.6 Layout of the GSIM Statistical Classification Model

41. Section 2 gives an overview of the GSIM Statistical Classification Model object types, including a short description. The list is ordered according to an obvious and simple logic.

42. A simplified object graph in Section 3 gives an overview of the main object types and relationships in the conceptual model.

43. Section 4 contains the list of all object types and their attributes. 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 attributes associated with the object type. Each attribute is also described. There has been an attempt also to order the attributes according to some sort of logic and to list them in a consistent way across the object types. Attributes or terms used in the descriptions which are underlined, refer to an object type listed and described elsewhere in the model. While object type terms are unique, the name of an attribute may differ in meaning when the attribute is associated with different object types. Some of the central object types of the model, e.g. Statistical Classification, Classification Item, 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 allowed 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 genuine definitions, sometimes indicating 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.

44. A worked example for all object types and most attributes, based mainly on the Standard Industrial Classification (SIC 2007), has been added in Appendix 1 to facilitate understanding. In Appendix 2, a checklist of possible content for the introduction to a classification version can be found.

1.7 References

- a) ISO 704: 2000. *Terminology Work - Principles and Methods*.
- b) *Terminology on Statistical Metadata*. Prepared by the UNECE Work Session on Statistical Metadata (METIS). Conference of European Statisticians, Statistical Standards and Studies No 53. Geneva 2000.
- c) *UN Glossary of Classification Terms*. Working document. United Nations. http://unstats.un.org/unsd/class/family/glossary_short
- d) Ward, D. and Pellegrino, M. *Developing a Common Understanding of Standard Metadata Components: A Statistical Glossary (draft)*. Joint OECD and Eurostat paper

for the Workshop on Statistical Data and Metadata Exchange, Washington , D.C., September 2001.

- e) *Generic Statistical Information Model*, UNECE,
<http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model>

2 Object types

45. The object types are more extensively described in the next section, which contains the main list of object types and their attributes.

Classification Family: A Classification Family is a group of Classification Series related from a particular point of view. The Classification Family is related by being based on a common concept (e.g. economic activity).

Classification Series: A Classification Series is an ensemble of one or several consecutive Statistical Classifications under a particular heading (for example ISIC or ISCO).

Statistical Classification: A Statistical Classification is a set of categories which may be assigned to one or more variables registered in statistical surveys or administrative files, and used in the production and dissemination of statistics.

Level: A Statistical Classification has a structure which is composed of one or several Levels. A Level often is associated with a concept, which defines it. In a hierarchical classification the Classification Items of each Level but the highest are aggregated to the nearest higher Level. A linear Statistical Classification has only one Level.

Correspondence Table: A Correspondence Table expresses the relationship between two Statistical Classifications.

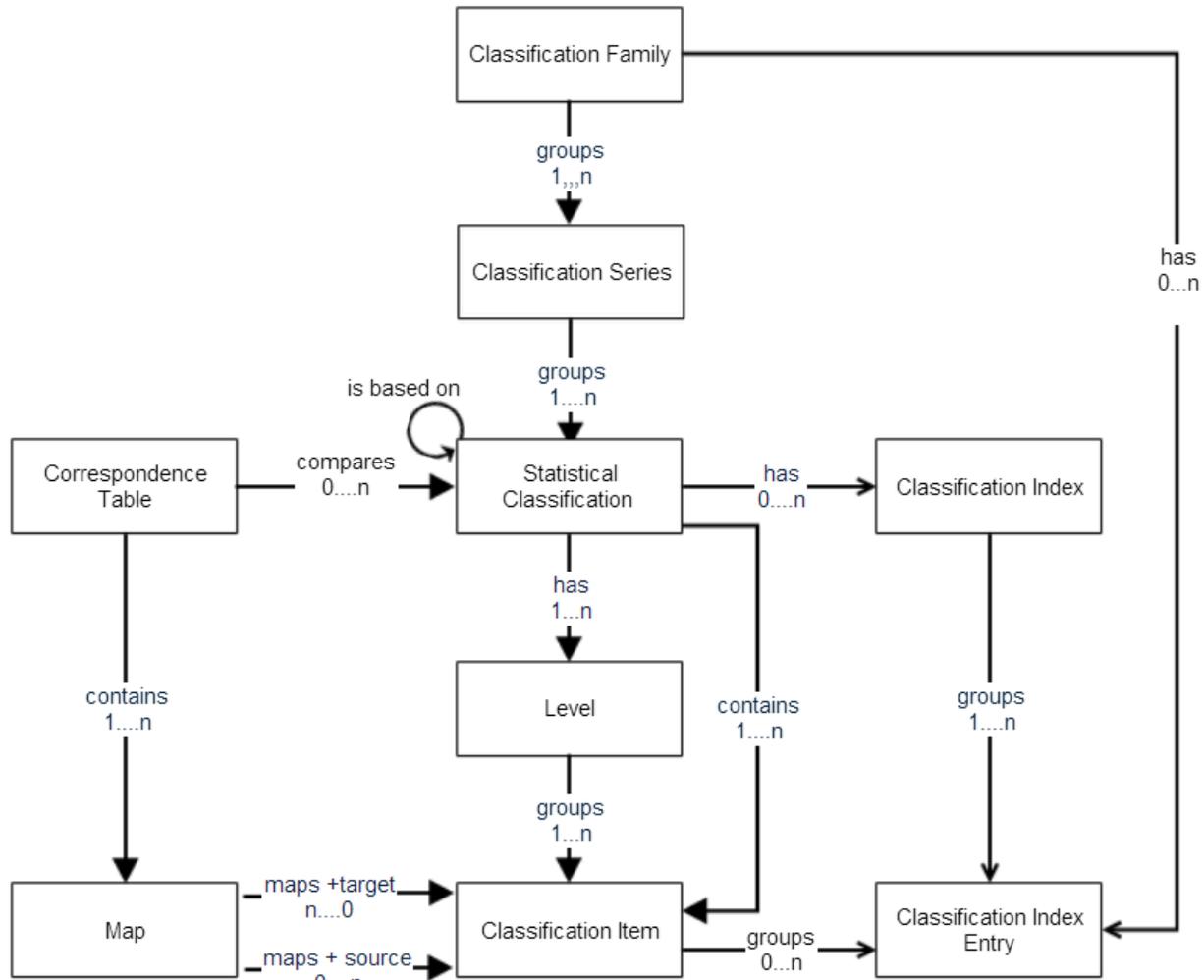
Classification Index: A Classification Index is an ordered list (alphabetical, in code order etc) of Classification Index Entries. A Classification Index relates to one particular or to several Statistical Classifications.

Classification Item: A Classification Item represents a Category at a certain Level within a Statistical Classification. It defines the content and the borders of the category. An object/unit can be classified to one and only one Classification Item at each Level of a Statistical Classification.

Map: An expression of the relation between a Classification Item in a source Statistical Classification and a corresponding Classification Item in the target Statistical Classification.

Classification Index Entry: A Classification Index Entry is a word or a short text (e.g. the name of a locality, an economic activity or an occupational title) describing a type of object/unit or object property to which a Classification Item applies, together with the code of the corresponding Classification Item.

3 Object graph



4 Object types and attributes

4.1 Classification Family

46. A Classification Family is a group of Classification Series related from a particular point of view. The Classification Family is related by being based on a common Concept (e.g. economic activity).

47. Different classification databases may use different types of Classification Families and have different names for the families, as no standard has been agreed upon.

Identifier: A Classification Family is identified by a unique identifier.

Name: A Classification Family has a name.

Classification Series: A Classification Family may refer to a number of Classification Series.

See also: Classification Series

4.2 Classification Series

48. A Classification Series is an ensemble of one or more Statistical Classifications, based on the same concept, and related to each other as versions or updates. Typically, these Statistical Classifications have the same name (for example, ISIC or ISCO).

Identifier: A Classification Series is identified by a unique identifier, which may typically be an abbreviation of its name.

Name: A Classification Series has a name as provided by the owner.

Description: Short general description of the Classification Series, including its purpose, its main subject areas etc.

Context: A Classification Series can be designed in a specific context.

Objects/units classified: A Classification Series is designed to classify a specific type of object/unit according to a specific attribute.

Subject areas: Areas of statistics in which the Classification Series is implemented.

Owners: The statistical office or other authority, which created and maintains the Statistical Classification (s) related to the Classification Series. A Classification Series may have several owners.

Keywords: A Classification Series can be associated with one or a number of keywords.

Classification Family: Classification Series may be grouped into Classification

Families. Shows to which Classification Family the Classification Series belongs.

Statistical Classification: A Classification Series has at least one Statistical Classification.

Current Statistical Classification: If there are several Statistical Classifications related to a Classification Series, one Statistical Classification *may be* assigned as the currently valid Statistical Classification.

See also: Statistical Classification, Classification Family

4.3 Statistical Classification

49. A Statistical Classification is a set of categories which may be assigned to one or more variables registered in statistical surveys or administrative files, and used in the production and dissemination of statistics. The categories at each level of the classification structure must be mutually exclusive and jointly exhaustive of all objects/units in the population of interest.

50. The categories are defined with reference to one or more characteristics of a particular population of units of observation. A Statistical Classification may have a flat, linear structure or may be hierarchically structured, such that all categories at lower Levels are sub-categories of categories at the next Level up. Categories in Statistical Classifications are represented in the information model as Classification Items.

Identifier: A Statistical Classification is identified by a unique identifier. The identifier of a Statistical Classification principally considered to be a version or update is typically an abbreviation of its name. It is often distinguished from other versions/updates of the same Classification Series by reference to a revision number or to the year when it came into force. The identifier of a Statistical Classification that is considered to be a variant typically refers to (contains) the identifier of its base Statistical Classification.

Name: A Statistical Classification has a name as provided by the owner or maintenance unit.

Introduction: The introduction provides a detailed description of the Statistical Classification, the background for its creation, the classification variable and objects/units classified, classification rules etc. See Appendix 2 for a checklist of possible topics to be included in the introduction.

Release date: Date on which the Statistical Classification was released.

Termination date: Date on which the Statistical Classification was superseded by a successor version or otherwise ceased to be valid.

Current: Indicates whether or not the Statistical Classification is currently valid.

Maintenance unit: The unit or group of persons within the organisation who are responsible for the Statistical Classification (i.e, for maintaining, updating and changing it).

Contact persons: Person(s) who may be contacted for additional information about the Statistical Classification.

Legal base: Indicates that the Statistical Classification is covered by a legal act or by some other formal agreement.

Publications: A list of the publications, including print, PDF, HTML and other electronic formats, in which the Statistical Classification has been published.

Name types: A list of the defined types of alternative item names available for the Statistical Classification. Each name type refers to a list of alternative item name.

Languages available: A Statistical Classification can exist in one or several languages. Indicates the languages available, whether the version is completely or partially translated, and which part is available in which language.

Copyright: Statistical Classifications may have restricted copyrights. Such Statistical Classifications might be excluded from downloading. Notes the copyright statement that should be displayed in official publications to indicate the copyright owner.

Dissemination allowed: Indicates whether or not the Statistical Classification may be published or otherwise disseminated (e.g. electronic dissemination).

Classification Series: A Statistical Classification is a version or update of one specific Classification Series.

Levels: The structure of a Statistical Classification is defined by its Levels (classification level). Include here links to the relevant Levels.

Items: A Statistical Classification is composed of categories structured in one or more Levels. Each category is represented by a Classification Item, which defines the content and the borders of the category.

Correspondence Tables: A Statistical Classification may be linked to other classification versions or classification variants through Correspondence Tables. Include here links to any relevant Correspondence Tables.

Classification Indexes: A Statistical Classification can be associated with one or a number of Classification Indexes in which Classification Index Entries are linked to the appropriate Classification Item. Include here links to any relevant Classification Indexes.

Version: Indicates if the Statistical Classification is a version.

Update: Indicates if the Statistical Classification is an update.

Floating: Indicates if the Statistical Classification is a floating classification. In a floating statistical classification, a validity period should be defined for all Classification Items which will allow the display of the item structure and content at different points of time.

Predecessor: For those Statistical Classifications that are versions or updates, notes the preceding Statistical Classification of which the actual Statistical Classification is the successor.

Successor: Notes the Statistical Classification that superceded the actual Statistical Classification.

Derived from: A Statistical Classification can be derived from one of the classification versions of another Classification Series. The derived Statistical Classification can either inherit the structure of the classification version from which it is derived, usually adding more detail, or use a large part of its Classification Items, rearranging them in a different structure. Indicates the classification version from which the actual Statistical Classification is derived.

Changes from previous version or update: A summary description of the nature and content of changes from the preceding version or update. Specific changes are recorded in the Classification Item object under the “Changes from previous version and updates” attribute.

Updates possible: Indicates whether or not updates are allowed within the classification version i.e. without leading to a new version. Indicate here what structural changes, if any, are permissible within a version. Note whether Classification Items can be added to the structure and whether they can be revalidated or invalidated. Such changes are more likely to be permissible in floating classifications. Also indicate whether changes to such things as Classification Item names and explanatory notes that do not involve structural changes are permissible within a version.

Updates: Summary description of changes which have occurred since the most recent classification version or classification update came into force.

Variants available: Identifies any variants associated with this version.

Variant: For those Statistical Classifications that are variants, notes the Statistical Classification on which it is based and any subsequent versions of that Statistical Classification to which it is also applicable.

Changes from base Statistical Classification: Describes the relationship between the variant and its base Statistical Classification, including regroupings, aggregations added and extensions.

Purpose of variant: If the Statistical Classification is a variant, notes the specific purpose for which it was developed.

See also: Classification Series, Level, Classification Item, Correspondence Tables, Classification Index.

4.4 Level

51. A Statistical Classification has a structure which is composed of one or several Levels. A Level often is associated with a concept, which defines it. In a hierarchical Statistical Classification the Classification Items of each Level but the highest are aggregated to the nearest higher Level. A linear Statistical Classification has only one Level.

Identifier: A Level is identified by a unique identifier.

Level number: The number associated with the Level. Levels are numbered consecutively starting with Level 1 at the highest (most aggregated) Level.

Level name: The name given to the Level.

Description: Text describing the content and particular purpose of the Level.

Number of Classification Items: The number of Classification Items (categories) at the Level.

Code type: Indicates whether the code at the Level is alphabetical, numerical or alphanumerical.

Code structure: Indicates how the code at the Level is constructed of numbers, letters and separators.

Dummy code: Rule for the construction of dummy codes from the codes of the next higher Level (used when one or several categories are the same in two consecutive Levels).

Items: An ordered list of the categories (Classification Items) that constitute the Level.

See also: Statistical Classification, Classification Item

4.5 Correspondence Table

52. A Correspondence Table expresses the relationship between two Statistical Classifications. These are typically: two versions from the same Classification Series; Statistical Classifications from different Classification Series; a variant and the version on which it is based; different versions of a variant. In the first and last examples, the Correspondence Table facilitates comparability over time. Correspondence relationships are shown in both directions.

Identifier: A Correspondence Table is identified by a unique identifier, which may typically include the identifiers of the versions or variants involved.

Name: A Correspondence Table has a name as provided by the owner.

Description: The description contains information about the scope and aim of the Correspondence Table and the principles on which it is based.

Owners: The statistical office, other authority or section that created and maintains the Correspondence Table. A Correspondence Table may have several owners.

Maintenance unit: The unit or group of persons who are responsible for the Correspondence Table, i.e. for maintaining and updating it.

Contact persons: The person(s) who may be contacted for additional information about the Correspondence Table.

Publications: A list of the publications in which the Correspondence Table has been published.

Source: The Statistical Classification from which the correspondence is made.

Target: The Statistical Classification(s) to which the correspondence is directed. There may be multiple target Statistical Classifications associated with the Correspondence Table.

Source level: The correspondence is normally restricted to a certain Level in the source Statistical Classification. In this case, target Classification Items are assigned only to source Classification Items on the given level. If no level is indicated, target Classification Items can be assigned to any Level of the source Statistical Classification.

Target level: The correspondence is normally restricted to a certain Level in the target Statistical Classification. In this case, source Classification Items are assigned only to target Classification Items on the given Level. If no Level is indicated, source Classification Items can be assigned to any Level of the target Statistical Classification.

Relationship type: A Correspondence Table can define a 1:1, 1:N, N:1 or M:N relationship between source and target Classification Items.

Floating: If the source and/or target Statistical Classifications of a Correspondence Table are floating Statistical Classifications, the date of the Correspondence Table must be noted. The Correspondence Table expresses the relationships between the two Statistical Classifications as they existed on the date specified in the Correspondence Table.

4.6 Classification Index

53. A Classification Index is an ordered list (alphabetical, in code order etc) of Classification Index Entries. A Classification Index can relate to one particular or to several Statistical Classifications.

54. A Classification Index shows the relationship between text found in statistical data sources (responses to survey questionnaires, administrative records) and one or more Statistical Classifications. A Classification Index may be used to assign the codes for Classification Items to observations in statistical collections.

Identifier: A Classification Index is identified by a name. If there are several Classification Indexes in different languages, the language should be part of the Classification Index name. If there are several Classification Indexes for different purposes, the purpose should be part of the Classification Index name. If there are several Classification Indexes that differ by language and by purpose, the language and the purpose should be part of the Classification Index name.

Release date: Date when the current version of the Classification Index was released.

Maintenance unit: The unit or group of persons within the organisation responsible for the Classification Index, i.e. for adding, changing or deleting Classification Index Entries.

Contact persons: Person(s) who may be contacted for additional information about the Classification Index.

Publications: A list of the publications in which the Classification Index has been published.

Languages: A Classification Index can exist in several languages. Indicates the languages available. If a Classification Index exists in several languages, the number of entries in each language may be different, as the number of terms describing the same phenomenon can change from one language to another. However, the same phenomena should be described in each language.

Corrections: Summary description of corrections, which have occurred within the Classification Index. Corrections include changing the item code associated with a Classification Index Entry.

Coding Instructions: Additional information which drives the coding process for all entries in a Classification Index.

Statistical Classification: A Classification Index is related to one particular Statistical Classification.

See also: Classification Index Entry, Statistical Classification

4.7 Classification Item

55. A Classification Item represents a category at a certain Level within a Statistical Classification. It defines the content and the borders of the category. An object/unit can be classified to one and only one Classification Item at each Level of a Statistical Classification.

Code: A Classification Item is identified by an alphabetical, numerical or alphanumerical code, which is in line with the code structure of the Level. The code is unique within the Statistical Classification to which the Classification Item belongs.

Official name: A Classification Item has a name as provided by the owner or maintenance unit. The name describes the content of the Category. The name is unique within the Statistical Classification to which the Classification Item belongs, except for Categories that are identical at more than one Level in a hierarchical Statistical Classification.

Alternative names: A Classification Item can be expressed in terms of one or several alternative names. Each alternative name is associated with a name type.

Explanatory notes: A Classification Item may be associated with explanatory notes, which further describe and clarify the contents of the category. Explanatory notes consist of:

General note: Contains either additional information about the category, or a general description of the category, which is not structured according to the "includes", "includes also", "excludes" pattern.

Includes: Specifies the contents of the category.

Includes also: A list of borderline cases, which belong to the described category.

Excludes: A list of borderline cases, which do not belong to the described category. Excluded cases may contain a reference to the Classification Items to which the excluded cases belong.

Level number: The number of the Level to which the Classification Item belongs.

Generated: Indicates whether or not the Classification Item has been generated to make the Level to which it belongs complete.

Currently valid: If updates are allowed in the Statistical Classification, a Classification Item may be restricted in its validity, i.e. it may become valid or invalid after the Statistical Classification has been released. Indicates whether or not the Classification Item is currently valid.

Valid from: Date from which the Classification Item became valid. The date must be defined if the Classification Item belongs to a floating Statistical Classification.

Valid to: Date at which the Classification Item became invalid. The date must be defined if the Classification Item belongs to a floating Statistical Classification and is no longer valid.

Future events: The future events describe a change (or a number of changes) related to an invalid Classification Item. These changes may e.g. have turned the now invalid Classification Item into one or several successor Classification Items. In describing

these changes, terminology from the Typology of item changes, found in Appendix 3 should be used. This allows the possibility to follow successors of the Classification Item in the future.

Changes from previous version or update: Describes the changes, which the Classification Item has been subject to from the previous version to the actual Statistical Classification. In describing these changes, terminology from the Typology of item changes, found in Appendix 3 should be used.

Updates: Describes the changes, which the Classification Item has been subject to during the life time of the actual Statistical Classification.

Statistical Classification: The Statistical Classification to which the Classification Item belongs.

Parent item: The Classification Item at the next higher Level of the Statistical Classification of which the actual Classification Item is a sub item.

Sub items: Each Classification Item, which is not at the lowest Level of the Statistical Classification, might contain one or a number of sub items, i.e. Classification Items at the next lower Level of the Statistical Classification.

Linked items: Classification Items of other Statistical Classifications with which the Classification Item is linked, either as source or target, through Correspondence Tables.

Case laws: Refers to identifiers of one or more case law rulings related to the Classification Item.

Case law descriptions: Refers to descriptions of the above case laws

Case law dates: Refers to dates of above case laws

See also: Level, Statistical Classification, Classification Index Entry

4.8 Map

56. A Map is an expression of the relation between a Classification Item in a source Statistical Classification and a corresponding Classification Item in the target Statistical Classification. The Map should specify whether the relationship between the two Classification Items is partial or complete. Depending on the relationship type of the Correspondence Table, there may be several Maps for a single source or target Classification Item.

Source item: The source item refers to the Classification Item in the source Statistical Classification.

Target item: The target item refers to the Classification Item in the target Statistical Classification.

Partial/complete: Specifies whether the relationship between the two Classification Items is partial or complete

Valid from: Date from which the Map became valid. The date must be defined if the Map belongs to a floating Correspondence Table.

Valid to: Date at which the Map became invalid. The date must be defined if the Map belongs to a floating Correspondence Table and is no longer valid.

See also: Statistical Classification, Classification Item, Correspondence Table

4.9 Classification index entry

57. A Classification Index Entry is a word or a short text (e.g. the name of a locality, an economic activity or an occupational title) describing a type of object/unit or object property to which a Classification Item applies, together with the code of the corresponding Classification Item. Each Classification Index Entry typically refers to one item of the Statistical Classification. Although a Classification Index Entry may be associated with a Classification Item at any Level of a Statistical Classification, Classification Index Entries are normally associated with Classification Items at the lowest Level.

Text: Text describing the type of object/unit or object property.

Statistical Classification: Identify the Statistical Classification(s) to which the Classification Index Entry is associated.

Codes: For each Statistical Classification to which the Classification Index Entry is associated, enter the code of the Classification Item in that Statistical Classification with which the Classification Index Entry is associated.

Valid from: Date from which the Classification Index Entry became valid. The date must be defined if the Classification Index Entry belongs to a floating Classification Index.

Valid to: Date at which the Classification Index Entry became invalid. The date must be defined if the Classification Index Entry belongs to a floating Classification Index and is no longer valid.

Coding Instructions: Additional information which drives the coding process. Required when coding is dependent upon one or many other factors.

See also: Classification Item, Classification Index, Statistical Classification

Appendix 1: Worked example of the GSIM Statistical Classification Model

58. In this appendix, a worked example for all object types and most attributes, based mainly on the Standard Industrial Classification (SIC 2007), is provided to facilitate understanding.

59. Attributes or terms used in the descriptions which are underlined, refer to an object type listed and described elsewhere in the model.

Classification Family

60. A Classification Family is a group of Classification Series related from a particular point of view. The Classification Family is related by being based on a common Concept (e.g. economic activity).

61. Different classification databases may use different types of Classification Families and have different names for the families, as no standard has been agreed upon.

Identifier: A Classification Family is identified by a unique identifier.

E.g.: IA

Name: A Classification Family has a name.

E.g.: Industrial activities

Classification Series: A Classification Family may refer to a number of Classification Series.

E.g.: Standard Industrial Classification, Classification of CPA codes

See also: Classification Series

62. For a practical example, see: <http://stabas.ssb.no/MainFrames.asp?Language=en>

Classification Series

63. A Classification Series is an ensemble of one or more Statistical Classifications, based on the same concept, and related to each other as versions or updates. Typically, these Statistical Classifications have the same name (for example ISIC or ISCO).

Identifier: A Classification Series is identified by a unique identifier, which may typically be an abbreviation of its name.

E.g.: SIC

Name: A Classification Series has a name as provided by the owner.

E.g.: Standard Industrial Classification

Description: Short general description of the Classification Series, including its purpose, its main subject areas etc.

E.g.: SIC is primarily a statistical standard. The standard will be the basis for coding units according to principal activity in e.g. a business register. The SIC is one of the most important standards of economic statistics, and it will make it possible to

compare and analyze statistical data both at the national/international level and over time. SIC is also used for administrative purposes.

Context: A Classification Series can be designed in a specific context.

E.g.: Not relevant

Objects/units classified: A Classification Series is designed to classify a specific type of object/unit according to a specific attribute.

E.g.: Economic activities

Subject areas: Areas of statistics in which the Classification Series is implemented.

E.g.: National accounts, Energy and manufacturing

Owners: The statistical office or other authority, which created and maintains the Statistical Classification(s) related to the Classification Series. A Classification Series may have several owners.

E.g.: Statistics Norway (The national version of SIC)

Keywords: A Classification Series can be associated with one or a number of keywords.

E.g.: Industry, business, legal units

Classification Family: Classification Series may be grouped into Classification Families. Shows which Classification Family the Classification Series belongs.

E.g.: Industrial activities

Statistical Classification: A Classification Series has at least one Statistical Classification.

E.g.: SIC94, SIC 2002, SIC 2007

Current Statistical Classification: If there are several Statistical Classifications related to a Classification Series, one Statistical Classification *may be* assigned as the currently valid Statistical Classification.

E.g.: SIC 2007

64. For a practical example, see:

<http://www.ssb.no/metadatas/classification/stabas/342101/en>

Statistical Classification

65. A Statistical Classification is a set of categories which may be assigned to one or more variables registered in statistical surveys or administrative files, and used in the production and dissemination of statistics. The categories at each level of the classification structure must be mutually exclusive and jointly exhaustive of all objects/units in the population of interest.

66. The categories are defined with reference to one or more characteristics of a particular population of units of observation. A Statistical Classification may have a flat, linear structure or may be hierarchically structured, such that all categories at lower Levels are sub-

categories of categories at the next Level up. Categories in Statistical Classifications are represented in the information model as Classification Items.

Identifier: A Statistical Classification is identified by a unique identifier. The identifier of a Statistical Classification principally considered to be a version or update is typically an abbreviation of its name. It is often distinguished from other versions/updates of the same Classification Series by reference to a revision number or to the year when it came into force. The identifier of a Statistical Classification that is considered to be variant typically refers to (contains) the identifier of its base Statistical Classification.

E.g.: SIC2007

Name: A Statistical Classification has a name as provided by the owner or maintenance unit.

E.g.: Standard Industrial Classification (SIC 2007)

Introduction: The introduction provides a detailed description of the Statistical Classification, the background for its creation, the classification variable and objects/units classified, classification rules etc. See Appendix 2 for a checklist of possible topics to be included in the introduction.

E.g.: Standard Industrial Classification is primarily a statistical standard. In practice, this means the standard will be the basis for coding units according to the most important activities in Statistics Norway's Business register and in the Central Coordinating Register for Legal Entities. SIC2007 is one of the most important standards of economic statistics, and it makes it possible to compare and analyze statistical data at the national/international level and over time.

Release date: Date on which the Statistical Classification was released.

E.g.: 01.01.2009

Termination date: Date on which the Statistical Classification was superseded by a successor version or otherwise ceased to be valid.

Current: Indicates whether or not the Statistical Classification is currently valid.

E.g.: Yes

Maintenance unit: The unit or group of persons within the organisation who are responsible for the Statistical Classification (i.e. for maintaining, updating and changing it).

E.g.: 810 - Division for statistical populations

Contact persons: Person(s) who may be contacted for additional information about the Statistical Classification.

E.g.: Ida Skogvoll, isk@ssb.no

Legal base: Indicates that the Statistical Classification version is covered by a legal act or by some other formal agreement.

E.g.: Council Regulation (EEC) No. 1893/2006

Publications: A list of the publications, including print, PDF, HTML and other electronic formats, in which the Statistical Classification has been published.
E.g.: http://www.ssb.no/a/publikasjoner/pdf/nos_d383/nos_d383.pdf

Name types: A list of the defined types of alternative item names available for the Statistical Classification. Each name type refers to a list of alternative item names.
E.g.: Short titles (for use in our dissemination tables)

Languages available: A Statistical Classification can exist in one or several languages. Indicates the languages available, whether the version is completely or partially translated, and which part is available in which language.
E.g.: Norwegian (bokmål), Norwegian (nynorsk), English

Copyright: Statistical Classifications may have restricted copyrights. Such Statistical Classifications might be excluded from downloading. Notes the copyright statement that should be displayed in official publications to indicate the copyright owner.
E.g.: Not relevant

Dissemination allowed: Indicates whether or not the Statistical Classification may be published or otherwise disseminated (e.g. electronic dissemination).
E.g.: Yes

Classification Series: A Statistical Classification is a version or update of one specific Classification Series
E.g.: Standard Industrial Classification

Levels: The structure of a Statistical Classification is defined by its Levels (classification levels). Include here links to the relevant Levels.
E.g.: Section, Division, Group, Class, Subclass

Classification Items: A Statistical Classification is composed of categories structured in one or more Levels. Each category is represented by a Classification Item, which defines the content and the borders of the category.
E.g.: 01.24 Growing of pome fruits and stone fruits

Correspondence Tables: A Statistical Classification may be linked to other classification versions or classification variants through Correspondence Tables. Include here links to any relevant Correspondence Tables.
E.g.: Correspondence table SIC 2007/SIC 2002

Classification Indexes: A Statistical Classification can be associated with one or a number of Classification Indexes in which Index Entries are linked to the appropriate Classification Item. Include here links to any relevant Classification Indexes.
E.g.: SN07XTD_en

Version: Indicates if the Statistical Classification is a version.
E.g.: Yes

Update: Indicates if the Statistical Classification is an update.

Floating: Indicates if the Statistical Classification is a floating Statistical Classification. In a floating Statistical Classification, a validity period should be defined for all Classification Items which will allow the display of the item structure and content at different points of time.

E.g.: No

Predecessor: For those Statistical Classifications that are versions or updates, notes the preceding Statistical Classification of which the actual Statistical Classification is the successor. .

E.g.: SIC 2002

Successor: Notes the Statistical Classification that superseded the actual Statistical Classification.

Derived from: A Statistical Classification can be derived from one of the classification versions of another Classification Series. The derived Statistical Classification can either inherit the structure of the classification version from which it is derived, usually adding more detail, or use a large part of its Classification Items, rearranging them in a different structure. Indicates the classification version from which the actual Statistical Classification is derived.

E.g.: NACE Rev.2

Changes from previous version or update: A summary description of the nature and content of changes from the previous version or update. Specific changes are recorded in the Classification Item object under the “changes from previous version and update” attribute.

E.g.: Information presented in **Publications** and in Correspondence Table

Updates possible: Indicates whether or not updates are allowed within the classification version, i.e. without leading to a new version. Indicate here what structural changes, if any, are permissible within the version. Note whether Classification Items can be added to the structure and whether they can be revalidated or invalidated. Such changes are more likely to be permissible in floating classifications. Also indicate whether changes to such things as Classification Item names and explanatory notes that do not involve structural changes are permissible within the version.

E.g.: No

Updates: Summary description of changes which have occurred since the most recent classification version or classification update came into force.

E.g.: Not relevant:

Variants available: Identifies any variants associated with this version.

E.g.: Variant of SIC - Environmental accounts (SIC2007)

Variant: For those Statistical Classifications that are variants, notes the Statistical Classification on which it is based and any subsequent versions of that Statistical Classification to which it is also applicable.

Changes from base Statistical Classification: Describes the relationship between the variant and its base Statistical Classification, including regroupings, aggregations added and extensions.

Purpose of variant: If the Statistical Classification is a variant, notes the specific purpose for which it was developed.

See also: Classification Series, Level, Classification Item, Correspondence Tables, Classification Index.

67. For a practical example, see:

<http://www.ssb.no/metadata/classversion/stabas/8118001/en>

Level

68. A Statistical Classification has a structure which is composed of one or several Levels. A Level often is associated with a Concept, which defines it. In a hierarchical Statistical Classification the Classification Items of each Level but the highest are aggregated to the nearest higher Level. A linear Statistical Classification has only one Level.

Identifier: A Level is identified by a unique identifier.

E.g.: SIC2007L5

Level number: The number associated with the Level. Levels are numbered consecutively starting with Level 1 at the highest (most aggregated) Level.

E.g.: 5

Level name: The name given to the Level.

E.g.: Subclass

Description: Text describing the content and particular purpose of the Level.

E.g.: Subclass is the most detailed level and describes the national level in SIC.

Number of Classification Items: The number of Classification Items (Categories) at the Level.

E.g.: 817

Code type: Indicates whether the code at the Level is alphabetical, numerical or alphanumerical.

E.g.: Numerical

Code structure: Indicates how the code at the Level is constructed of numbers, letters and separators.

E.g.: nn.nnn

Dummy code: Rule for the construction of dummy codes from the codes of the next higher Level (used when one or several Categories are the same in two consecutive Levels).

E.g.: Not relevant

Items: An ordered list of the Categories (Classification Items) that constitute the Level.

E.g.: <http://www.ssb.no/metadata/codelist/stabas/8118013/en>

See also: Statistical Classification, Classification Item

Correspondence table

69. A Correspondence Table expresses the relationship between two Statistical Classifications. These are typically: two versions from the same Classification Series; Statistical Classifications from different Classification Series; a variant and the version on which it is based; different versions of a variant. In the first and last examples, the Correspondence Table facilitates comparability over time. Correspondence relationships are shown in both directions.

Identifier: A Correspondence Table is identified by a unique identifier, which may typically include the identifiers of the versions or variants involved.

E.g.: C20072002

Name: A Correspondence Table has a name as provided by the owner.

E.g.: SN2007, SN2002

Description: The description contains information about the scope and aim of the Correspondence Table and the principles on which it is based.

E.g.: The Correspondence Table shows the changes between SIC versions 2002 and 2007 and makes comparability over time possible.

Owners: The statistical office, other authority or section that created and maintains the Correspondence Table. A Correspondence Table may have several owners.

E.g.: Statistics Norway

Maintenance unit: The unit or group of persons who are responsible for the Correspondence Table, i.e. for maintaining and updating it.

E.g.: 810 - Division for statistical populations

Contact persons: The person(s) who may be contacted for additional information about the Correspondence Table.

E.g.: Ida Skogvoll, isk@ssb.no

Publications: A list of the publications in which the Correspondence Table has been published.

E.g.: The Correspondence Table is only published in the classification database

Source: The Statistical Classification from which the correspondence is made.

E.g.: SIC 2007

Target: The Statistical classification(s) to which the correspondence is directed. There may be multiple target Statistical Classifications associated with the Correspondence Table.

E.g.: SIC 2002

Source level: The correspondence is normally restricted to a certain Level in the source Statistical Classification. In this case target Classification Items are assigned only to source Classification Items on the given Level. If no Level is indicated target Classification Items can be assigned to any Level of the source Statistical Classification.

E.g.: Level 5

Target level: The correspondence is normally restricted to a certain Level in the target Statistical Classification. In this case source Classification Items are assigned only to target Classification Items on the given Level. If no Level is indicated, source Classification Items can be assigned to any Level of the target Statistical Classification.

E.g.: Level 5

Relationship type: A correspondence can define a 1:1, 1:N, N:1 or M:N relationship between source and target Classification Items.

E.g.: M:N

Floating: If the source and/or target Statistical Classifications of a Correspondence Table are floating Statistical Classifications, the date of the Correspondence must be noted. The Correspondence Table expresses the relationships between the two Statistical Classifications as they existed on the date specified in the Correspondence Table.

E.g.: No

See also Statistical Classification, Classification Item, Level, Map

70. For a practical example, see:

<http://stabas.ssb.no/CorrTabFrames.asp?ID=8364101&Language=en>

Classification index

71. A Classification Index is an ordered list (alphabetical, in code order etc.) of Classification Index Entries. A Classification Index can relate to one particular or to several Statistical Classifications. A Classification Index shows the relationship between text found in statistical data sources (responses to survey questionnaires, administrative records) and one or more statistical classifications. A Classification Index may be used to assign the codes for Classification Items to observations in statistical collections.

Identifier: A Classification Index is identified by a name. If there are several Classification Indexes in different languages, the language should be part of the Classification Index name. If there are several Classification Indexes for different purposes, the purpose should be part of the Classification Index name.. If there are several Classification Indexes that differ by languages and by purpose, the language and the purpose should be part of the Classification Index name.

E.g.: SN07XTD_en

Release date: Date when the current version of the Classification Index was released.

E.g.: 01.01.2009

Maintenance unit: The unit or group of persons within the organisation responsible for the Classification Index, i.e. for adding, changing or deleting Classification Index entries.

E.g.: 810 - Division for statistical populations

Contact persons: Person(s) who may be contacted for additional information about the Classification Index.

E.g.: Ida Skogvoll, isk@ssb.no

Publications: A list of the publications in which the classification index has been published.

Languages: A classification index can exist in several languages. Indicates the languages available. If an Classification Index exists in several languages, the number of entries in each language may be different, as the number of terms describing the same phenomenon can change from one language to another. However the same phenomena should be described in each language.

E.g.: Norwegian (bokmål), English

Corrections: Verbal summary description of corrections, which have occurred within the Classification Index. Corrections include changing the item code associated with an index entry.

Coding instructions: Additional information which drives the coding process for all entries in a Classification Index.

Statistical Classification: A Classification Index is related to one particular Statistical Classification.

E.g.: SIC 2007

Classification item

72. A classification item represents a Category at a certain Level within a Statistical Classification. It defines the content and the borders of the Category. A statistical object/unit can be classified to one and only one Classification Item at each level of a Statistical Classification.

Code: A Classification Item is identified by an alphabetical, numerical or alphanumerical code, which is in line with the code structure of the Level. The code is unique within the Statistical Classification to which the Classification Item belongs.

E.g.: 01.620

Official name: A Classification Item has a name as provided by the owner or maintenance unit. The name describes the content of the Category. The name is unique within the Statistical Classification to which the Classification Item belongs, except for Categories that are identical at more than one Level in a hierarchical Statistical Classification.

E.g.: Support activities for animal production

Alternative names: A Classification Item can be expressed in terms of one or several alternative names. Each alternative name is associated with a name type.

E.g.: Short text: Support activities to agriculture

Explanatory notes: A Classification Item may be associated with explanatory notes, which further describe and clarify the contents of the category. Explanatory notes consist of:

General note: Contains either additional information about the Category, or a general description of the Category, which is not structured according to the "includes", "includes also", "excludes" pattern.

E.g.: Not relevant

Includes: Specifies the contents of the Category.

E.g.: Includes services, associated with the keeping of farm animals, in activities that increase reproduction, growth and performance in farm animals, testing of farm animals (control), maintenance of grazing areas, castration, cleaning of barns, insemination and covering, clipping of sheep, housing and care of farm animals.

E.g.: Includes activities in connection with agriculture carried out on a contract basis

Includes also: A list of borderline cases, which belong to the described Category.

E.g.: Includes also: Includes also shoeing of horses.

Excludes: A list of borderline cases, which do not belong to the described Category. Excluded cases may contain a reference to the Classification Items to which the excluded cases belong.

E.g.: Excludes: Renting out of areas exclusively for housing of farm animals is grouped under Other letting of real estate. Veterinary services is grouped under 75.000 Veterinary activities, Vaccination of farm animals is grouped under 75.000 Veterinary activities. Renting out of farm animals (e.g. cattle) is grouped under 77.390 Renting and leasing of other machinery, equipment and tangible goods n.e.c. Housing of domestic pets is grouped under 96.09 Other personal service activities n.e.c.

Level number: The number of the Level to which the Classification Item belongs.

E.g.: 5

Generated: Indicates whether or not the Classification Item has been generated to make the Level to which it belongs complete.

E.g.: No

Currently valid: If updates are allowed in the Statistical Classification, a Classification Item may be restricted in its validity, i.e. it may become valid or invalid after the Statistical Classification has been released. Indicates whether or not the Classification Item is currently valid.

E.g.: Not relevant

Valid from: Date from which the Classification Item became valid. The date must be defined if the Classification Item belongs to a floating Statistical Classification.

E.g.:

Valid to: Date at which the Classification Item became invalid. The date must be defined if the Classification Item belongs to a floating Statistical Classification and is no longer valid.

E.g.:

Future events: The future events describe a change (or a number of changes) related to an invalid Classification Item. These changes may e.g. have turned the now invalid Classification Item into one or several successor Classification Items. In describing these changes, terminology from the Typology of item changes, found in Appendix (1) should be used. This allows the possibility to follow successors of the Classification Item in the future.

E.g.: Not relevant

Changes from previous version of the Statistical Classification: Describes the changes, which the Classification Item has been subject to from the previous to the actual Statistical Classification. In describing these changes, terminology from the Typology of item changes, found in Appendix (1), should be used.

E.g.: Name and Code change. In SIC2002 this item was called "01.420 Animal husbandry service activities, except veterinary activities". The history of the item is also documented in the Correspondence Table.

Updates: Describes the changes, which the Classification Item has been subject to during the life time of the actual Statistical Classification.

E.g.: Not relevant

Statistical Classification: The Statistical Classification to which the Classification Item belongs.

E.g.: SIC 2007

Parent item: The Classification Item at the next higher Level of the Statistical Classification of which the actual Classification Item is a sub item.

E.g.: 01.62 Support activities for animal production

Sub items: Each Classification Item, which is not at the lowest Level of the Statistical Classification, might contain one or a number of sub items, i.e. Classification Items at the next lower Level of the Statistical Classification.

E.g.: Not relevant

Linked items: Classification Items of other Statistical Classification with which the Classification Item is linked, either as source or target, through Correspondence Tables.

E.g.: 01.420 Animal husbandry service activities, except veterinary activities (SIC2002)

Case law: Refers to identifiers of one or more case law rulings related to the Classification Item.

E.g.: Not relevant

Case law descriptions: Case law descriptions refers to descriptions of the above case laws.

E.g.: Not relevant

Case law dates: Refers to dates of above case laws.

E.g.: Not relevant

See also Level, Classification Index Entry, Correspondence Item, Statistical Classification

Map

73. A map is an expression of the relation between a Classification Item in a source Statistical Classification and a corresponding Classification Item in the target Statistical Classification. The Map should specify whether the relationship between the two Classification Items is partial or complete. Depending on the relationship type of the Correspondence Table, there may be several maps for a single source or target Classification Item.

Source item: The source item refers to the Classification Item in the source Statistical Classification.

E.g.: 01.620 Support activities for animal production

Target item: The target item refers to the Classification Item in the target Statistical Classification.

E.g.: 01.420 Animal husbandry service activities, except veterinary activities

Partial/complete: specifies whether the relationship between the two Classification Items is partial or complete

E.g.: Complete

Valid from: Date from which the Map became valid. The date must be defined if the Map belongs to a floating Correspondence Table.

Valid to: Date at which the Map became invalid. The date must be defined if the Map belongs to a floating Correspondence Table and is no longer valid.

See also Statistical Classification, Classification Item, Correspondence Table

Classification Index Entry

74. A Classification Index Entry is a word or a short text (e.g. the name of a locality, an economic activity or an occupational title) describing a type of object/unit or object property to which a Classification Item applies, together with the code of the corresponding Classification Item. Each Classification Index Entry typically refers to one item of the Statistical Classification. Although a Classification Index Entry may be associated with a

Classification Item at any level of the Statistical Classification, Classification Index Entries are normally associated with Classification Items at the lowest Level.

Text: Text describing the type of object/unit or object property.

E.g.: Animal husbandry

Statistical Classification: Identifies the Statistical Classification(s) to which the Classification Index Entry is associated.

E.g.: SIC2007

Codes: For each Statistical Classification to which the Classification Index Entry is associated, enter the code of the Classification Item in that Statistical Classification with which the Classification Index Entry is associated.

E.g.: 01.620 Support activities for animal production

Valid from: Date from which the Classification Index Entry became valid. The date must be defined if the Classification Index Entry belongs to a floating Classification Index.

Valid to: Date at which the Classification Index Entry became invalid. The date must be defined if the Classification Index Entry belongs to a floating Classification Index and is no longer valid.

Coding instructions: Additional information which drives the coding process. Required when coding is dependent upon one or many other factors.

See also Classification Item, Classification Index, Statistical Classification

Appendix 2: Checklist of possible content for the Introduction to a Statistical Classification

Detailed description of the Statistical Classification

75. Describe the structure of the classification, including the number of Levels, their names, the number of Classification Items at each Level, the structure of the codes and the relationship between the codes used at the different Levels.

76. Describe the manner in which information is presented. What elements can the user find in the definitions of each Level of the Statistical Classification? Are all available examples presented in the main text or are some only in the index?

Background for the creation of this Statistical Classification

77. Identify the previous version of the Statistical Classification if relevant, as well as the reasons for the update. The revision process could be described, including any consultations that occurred.

Relationship of this Statistical Classification to other Statistical Classifications

78. Identify other Statistical Classifications that classify related subject matter and discuss how they relate to each other. For example, a classification of occupations might include in its introduction discussion of the relationship of this classification to classifications of industry or class of worker. The introduction to an industry classification might include its relationship to classifications of products.

Other Statistical Classifications applicable to the same subject matter

79. Identify other classifications that refer to the same subject matter in order to alert users to options available for coding and analysis.

Relationship of this Statistical Classification to relevant international standard classifications

80. Identify any related international standards and discuss: the degree to which this classification is coherent with the international standard; the nature of any differences; and, the reasons for such differences.

Summary of changes from the previous version

81. Summarize the changes from the previous version. The information provided here would not be as specific as that in the Correspondence Table but would summarize, for example, the number of new Classification Items at each Level, the number of Classification Items at each Level that were collapsed, the extent to which new Classification Index Entries have been added, the extent to which definitions or Designations have been revised, and identification of any particular sections of the previous Statistical Classification that were more extensively revised.

Classification criteria

82. Identify the criteria on which units have been grouped together in this Statistical Classification.

83. If different classification criteria are used, or given primacy, at different levels of the Statistical Classification, this should be discussed. For example, in Canada's National Occupational Classification, the main classification criteria are skill level and skill type. While both criteria apply at the unit group and minor group levels, the major group level is defined by skill type only.

84. If certain criteria apply only in specific parts of the classification, this should be discussed. For example, in Canada's National Occupational Classification, industry is used as a classification criterion but only in areas of the classification where it could be relevant to users, such as areas referring to industry-specific occupations where internal progression ladders are typical.

Objects / units classified

85. Identify here the nature of the Unit Type to which this classification can be applied. This may require definition of the Unit Type, such as "enterprise", and the specification of criteria for identifying a unit.

86. Include here discussion of the Unit Types to which the classification can be directly applied as well as those to which it is typically applied indirectly. For example, a classification of occupations classifies jobs. However, it is typically used to classify people on the basis of some job to which they are associated. This could be, for example, their current job, their most recent job or, for those with more than one job, the job at which they work the most hours per week.

The classification Concept

87. Identify the underlying Concept that is measured/described by this Statistical Classification. Define this Concept and provide any relevant clarifications regarding the scope of the coverage of the Concept provided by the classification. For example, the introduction of an occupational classification could clarify whether subsistence economic activity is included in its conceptual coverage and the extent to which unpaid activity, such as housework and child care, are included.

Information required for coding

88. Discuss here the information required about a particular Unit Type in order to classify it. For example, to classify a person's job it is necessary to have information about their job title and about the main duties they perform. It can also be helpful to know the industry in which the person is working, the level of education they have attained and their field of study.

How to use the Statistical Classification in coding

89. Describe how best to use the Statistical Classification to apply codes to particular Units. For example, discuss here whether, or under what circumstances, coders should start at the most highly aggregated Level and make a series of choices to work down the classification to find the best detailed Level code for a particular observed Unit.

90. Discuss the relative primacy to give to competing pieces of information, such as job title versus duties performed.

Classification rules

91. Discuss here any specific rules related to applying the Statistical Classification in coding. This can include, for example, how to code double responses to a question about occupation title or main field of study, or how to code enterprises that are engaged in multiple activities.

92. Describe any rules that apply to specific parts of the Statistical Classification. For example, the introduction for an occupational classification could describe rules for classifying to management occupations as opposed to supervisory occupations.

93. Discuss any rules or principles that have been developed to facilitate the implementation of specific Classification Items. For example, the introduction to an industry classification could discuss the conditions under which production units engaged in e-commerce are to be coded to sales.

Variants

94. Present the structure of any variants known at the time the Statistical Classification is published.

95. Show how they relate to the Statistical Classification and discuss the appropriate contexts in which each variant is to be used.

Appendix 3: A typology of Classification Item changes

1. Introduction

96. A Correspondence Table expresses the relationship between two Statistical Classifications. These relationships show how the items of the source/predecessor Statistical Classification relate to the items of the target/successor Statistical Classification. The typology presented in this appendix provides a systematic way to classify the various types of relationships, and thus enables to enrich the information held in Correspondence Tables.

97. Before presenting the typology, we have to make clear what we understand by “change of a Classification Item”.

2. Features of Item Change

98. A Classification Item is the designation of a category of a Statistical Classification. It can be said to consist of three components:

- a code
- a name (official title)
- a definition, which is:
 - a. expressed in explanatory notes, index entries and case law,
 - b. expressed in the set of statistical objects/units, belonging to the category. This is the *denotation* of the category.

99. From these three components, the definition represents the essence of the category best, more than the name and the code. Therefore, we will consider the *meaning* of a category to be represented by its definition, especially its denotation.

100. Ideally, naming, coding and definition ought to behave consistently, while the definition is “leading”, in the sense that a change of definition⁶ should necessarily imply a change of code and name, while conversely a code or name may only change if accompanied by a meaning change. In practice, however it may occur that a code or a name of an item changes while its definition remains the same, or the other way around. The typology should account for such situations. This is done by distinguishing between *real change* and *virtual change*. Where real change stands for changes in meaning, whether or not accompanied by changes in naming and/or coding, while virtual change stands for changes in coding and/or naming, while the meaning remains the same.

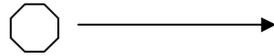
3. A Typology of Item Changes

101. The figures show the relationships between Classification Items involved in the change from the old (left) to the new (right) Statistical Classification. Blocks denote Classification Items. Classification Items are identified by their code. A block that keeps its colour keeps its code. The various changes are expressed in terms of events.

⁶ Note that not all changes in (the wording of) definitions imply changes in meaning. Indeed, changes which “merely” intend to enhance the *clarity* of a definition, or to further *operationalise* a definition on the basis of situations encountered in practice, do not affect the *meaning* of an item. The same goes for extending a Classification Index or case law.

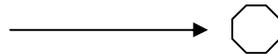
A. REAL CHANGE

A.1. Deletion (1 : 0)



A Classification Item expires, while its denotation reduces to zero; it does not proceed as (part of the) denotation of one or more other (existing) Classification Items.

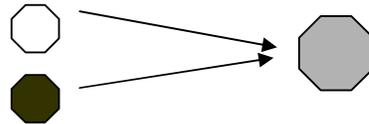
A.2 CREATION (0 : 1)



The mirror image of deletion: a Classification Item emerges, while its denotation is not (part of) the denotation of one or more existing Classification Items.

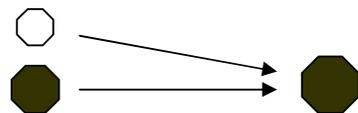
A.3 COMBINATION (N:1)

A.3.1 MERGER



Two or more Classification Items expire, while their denotations proceed in one emerging Classification Item.

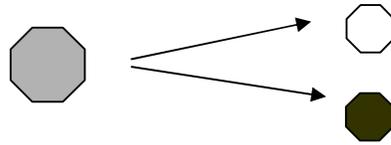
A.3.2 TAKE-OVER (N: 1)



A Classification Item expires, while its denotation proceeds as part of the denotation of another Classification Item, which continues its existence.

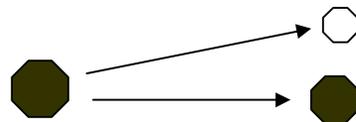
A.4 DECOMPOSITION (1: N)

A.4.1 BREAKDOWN



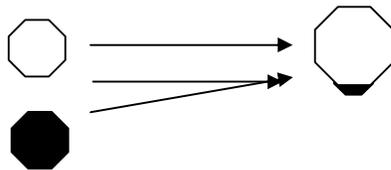
The mirror image of merger: a Classification Item expires, while its denotation is distributed over and proceeds in two or more emerging Classification Items.

A.4.2 SPLIT OFF



The mirror image of take-over: a Classification Item continues to exist, while part of its denotation moves to another (emerging) Classification Item.

A.5 TRANSFER (M : N)

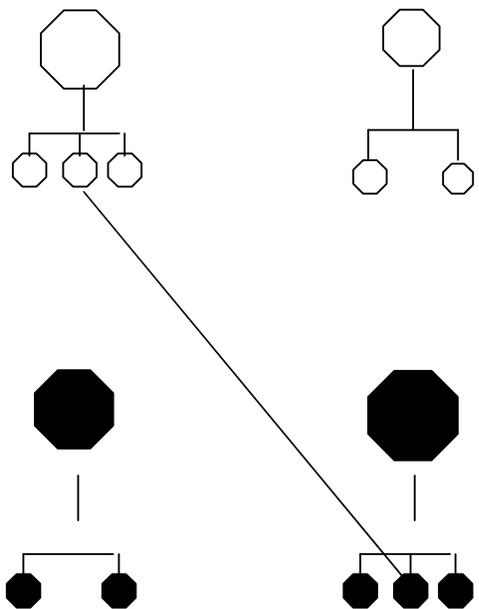


Part of the denotation of a Classification Item moves to another (existing) Classification Item

102. Notes:

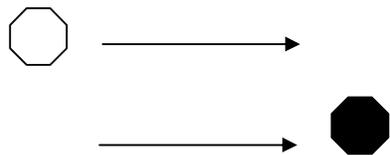
1. The situation pictured is the most simple of a number of situations, in which more Classification Items may be involved in the relationship between the old and new structure.

2. "Transfer" can also be applied at higher Levels of a Statistical Classification. In such cases *the part of the denotation that moves* (to another (existing) Classification Item) corresponds to a Classification Item at the lower Level:



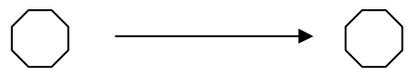
B. VIRTUAL CHANGE

B.1 CODE CHANGE⁷ (1:0;0:1)



A Classification Item expires, while its denotation proceeds as the denotation of an emerging Classification Item.

B.2 NAME CHANGE (1:1)



The name of a Classification Item changes, while its denotation remains the same.

103. Deletion and creation may result in border changes. Combination, decomposition and transfer do result in border changes. Code change and name change do not result in border changes.

⁷ Strictly speaking, this heading is not correct, because different codes denote different Classification Items.