

18th June 2004

Neuchâtel Terminology Model

Classification database object types and their attributes

Version 2.1

Statistics Denmark

Søren Netterstrøm (sne@dst.dk)

Statistics Norway

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

Statistics Sweden

Birgitta Ehrenström, editor (birgitta.ehrenstrom@scb.se)

Swiss Federal Statistical Office

Claude Macchi (claude.macchi@bfs.admin.ch)

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

Run Software-Werkstatt

Reinhard Karge (reinhard.karge@runsoftware.com)

Contents

Summary

1 Introduction

- 1.1 Background
- 1.2 Context and scope
- 1.3 Classification and related concepts
 - 1.3.1 Classification
 - 1.3.2 Classification and classification version, Neuchâtel-specific
 - 1.3.3 Classification variant
 - 1.3.4 Value set
- 1.4 Other terminologies
- 1.5 Implementation
- 1.6 Layout of the terminology
- 1.7 References

2 Object types

3 Object types and attributes

- 3.1 Classification family
- 3.2 Classification
- 3.3 Classification version
- 3.4 Classification variant
- 3.5 Classification index
- 3.6 Correspondence table
- 3.7 Classification level
- 3.8 Classification item
- 3.9 Case law
- 3.10 Classification index entry
- 3.11 Correspondence item

4 Object graph

Appendix 1: Implementation in Statistics Denmark and Statistics Norway

Appendix 2: Implementation in the Bridge^{NA} metadata software

Appendix 3: Types of Item Change

Neuchâtel Terminology

Classification database object types and their attributes

Summary

The Neuchâtel Terminology was developed by the Neuchâtel group in an endeavour to arrive at a common language and a common perception of the structure of classifications and the links between them. The Terminology is both a terminology and a conceptual model. It defines the key concepts that are relevant for how to structure classification metadata and provides the conceptual framework for the development of a classification database.

The Neuchâtel model has a two level structure, consisting at the first level of the object types of the database (e.g. classification, classification version, 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 a classification database. The Neuchâtel conceptual model is generally applicable and not dependent on IT software and platforms. It may be used in any context where structured information on classifications is required. A simplified object graph gives an overview of the main object types and relationships in the conceptual model.

Appendices 1 and 2 describe two different implementations of the Neuchâtel model. Appendix 3 gives a typology for item changes.

1 Introduction

1.1 Background

The Statistical Offices of Denmark, Norway and Sweden were among the participants of the IMIM project (Integrated Metainformation Management System) of the 4th Framework Programme of the EU. The main result of the IMIM project is a software product, BRIDGE, an object-oriented system for metadata management. In 1999, an informal co-operation between the statistical offices of Denmark, Norway, Sweden and Switzerland was started, focussing on the classification database part of BRIDGE. In June 1999, a meeting on terminology was held in Neuchâtel, Switzerland, with participants from the above Statistical Offices and run Software-Werkstatt (developers of the BRIDGE software). This was the birth of the "Neuchâtel Group"¹.

The aim of the Neuchâtel meeting was to clarify some basic concepts and to arrive at a common terminology for classifications. The meeting resulted in a draft terminology document and a

¹ The first Neuchâtel meeting was chaired by Heinrich Brünger, Swiss Federal Statistical Office. Although participation at the Neuchâtel group meetings has varied, the core of the group is made up of Birgitta Ehrenström, Statistics Sweden, Anne Gro Hustoft, Statistics Norway, Claude Macchi, Swiss Federal Statistical Office, Søren Netterstrøm, Statistics Denmark and Reinhard Karge, run Software-Werkstatt. The late Håkon Berby of Statistics Norway participated in the group from the beginning and greatly contributed to the development of the Neuchâtel Terminology. The result of this group's work was presented in the Neuchâtel Terminology version 2.0. In this version, 2.1, the paper has been somewhat adjusted and extended, and the group has been extended with the participants Nadja du Hamél from run Software-Werkstatt, Jenny Linnerud from Statistics Norway and Ad Willeboordse from Statistics Netherlands.

tentative conceptual model, in which the typical object types of a classification database, and the attributes connected with each object type, are listed and described.

The draft terminology and conceptual model was corrected, amended and further developed during 1999 and 2000. Version 1 was presented at the METIS meeting in Washington in November 2000 and at subsequent Eurostat or other international meetings on metadata. In this updated version, Version 2.0, some amendments and additions have been made and a few errors corrected. The introduction has been partly rewritten.

1.2 Context and scope

Classifications are generally regarded as a special kind of metadata for statistics. They are definitional, contents-oriented metadata, ordering and describing the meaning of statistical data. A classification database can be described as a register of metaobjects (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.

In spite of its name, the Neuchâtel terminology is not only a collection of terms used to designate concepts related to classifications. It has been developed for a particular purpose, using a specific methodology, ordering 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, correspondence table, index) and, on the second level, it lists the attributes connected with each object type. The Neuchâtel terminology 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 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 is to 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 Neuchâtel 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 a classification database.

1.3 Classification and related concepts

According to ISO 704: 1987 (E) Principles and methods of terminology, a term is a word or phrase, which designates a concept. The initial aim of the Neuchâtel meeting was to arrive at a common understanding of some central concepts and the words used to designate them. In the context at hand this implies having a common perception of the structure of classifications, of the components that make up classification metadata and the relationships between these components. However, agreeing on a concept does not necessarily entail agreeing on the word to designate it. Choosing which terms to use for the various concepts has not been a simple matter.

1.3.1 Classification

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).

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.

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 the Neuchâtel terminology.

1.3.2 Classification and classification version, Neuchâtel-specific

In the conceptual framework of the classification database, it was necessary to identify an object type roughly equivalent to concept d. in the preceding paragraph. In the Neuchâtel terminology it is this particular concept that has been named **classification**. The concept of each “structured list of mutually exclusive categories” has been named **classification version**. Thus, here, “classification” is not in itself a list of mutually exclusive categories. It is a kind of umbrella or “name” (e.g. NACE) that comprises or holds together one or several lists of mutually exclusive categories, i.e. classification versions, which are consecutive over time and closely related to each other (e.g. NACE 1970, NACE Rev.1).

In spite of this distinction, the term “classification” is used rather loosely in this document. It is probably sufficiently well understood. The distinction between classification and classification version is thus not upheld in a strict way in the introduction or in the descriptive parts of the terminology. Doing so would unnecessarily burden the texts with long and cumbersome names. The terms **classification level**, **classification item**, **classification index** have thus been used for concepts, which should rightly be named classification version level, classification version item, classification version index.

1.3.3 Classification variant

Classifications and their versions have some sort of official or normative status. Sometimes they are covered by a legal act or other formal agreement. It is assumed by the “owner” of the classification that the version will answer all needs and be implemented as it is. This is not always the case. Be it for dissemination or other uses, the classification structure may be found ill suited for the purpose at hand: Levels or categories are too general or too narrow, too detailed in one area, and too broad in another. Over the lifetime of a classification version, a number of variants may be created, in which the original categories are split or regrouped to provide context-specific additions or alternatives to the standard aggregation structure.

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.

It is yet another matter to decide whether a non-standard grouping based on a classification version is truly a variant of the version or not. The criteria for grouping may differ so much from the grouping criteria laid down for the base classification version that the actual non-standard grouping

should rather be treated as a new and different classification.

Concepts related to variants or non-standard groupings were introduced in the terminology from the start. They have appeared in earlier drafts in different shapes and under different names. When Version 1 of the terminology was to be made public, it was felt that the issue of variants had not been sufficiently explored. It was decided that the object types and attributes relating to variants should be removed until it became clearer how the concepts should be delineated. In Version 2.0, the object type Classification variant has been reintroduced. It has been given almost the same set of attributes as Classification version, even though some of these attributes may not often be applicable and only a limited number are likely to be used.

1.3.4 Value set

The management of value sets and variables does not normally form a part of a classification database. The Neuchâtel terminology does not deal with value sets. Yet a classification (or, to be specific, a classification version) is essentially a value set, in the sense that it contains the valid values for the classification variable. This is most obviously the case with a linear classification.

A classification is thus a particular kind of value set or grouping of value sets. What distinguishes it from any other value set is its more conceptual character and its normative status. The categories of a classification have a considerable stability over time and can only be changed through a more or less formalised revision process. Value sets in general are of a more temporary and operational character and are more immediately associated with the making of statistics. Of the variables used in a statistical organisation, only a limited number are classification variables, i.e. variables whose valid values correspond to the categories of a classification. The total number of value sets is greater by far than the number of classifications.

In spite of the above distinction, there are no strict rules to apply when trying to decide whether or not a particular value set (or value set grouping) should be included in the classification database. Not all classifications are covered by a legal act or other formal agreement. Some value sets are regarded as classifications simply because they have proved to be stable and are widely used. In the practice of building a classification database, where to draw the line between what to include and what to leave out will always be a matter of choice.

1.4 Other terminologies

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 Neuchâtel terminology. 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. There is also the UN/ECE "METIS" Terminology on Statistical Metadata. This has the term Classification scheme instead of Classification but the concept is the same.

The Eurostat Concepts and Definitions Database, CODED, is another tool for the harmonization of concepts. It is designed to hold concepts and definitions used in all areas of statistics covered by Eurostat. CODED lists a large number of variables, types of statistical objects and other terms relating to specific areas of statistics, as well as a number of classifications. The Neuchâtel terminology and CODED are situated at different conceptual levels and don't seem to have any concepts in common.

Concepts and terms related to classifications are also found in more general papers, as for instance the report Standard Statistical Classifications: Basic Principles, a paper presented to the 30th session of the *UN Statistical Commission* in March 1999. This paper describes best practices for the development, use, maintenance and revision of classifications. Both its usage and scope are different from those of the Neuchâtel terminology.

The Neuchâtel 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 between the Neuchâtel terminology and 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. Only the Neuchâtel terminology has an explicit listing of the different meanings of the term “classification”. Not surprisingly, the one instance in which the Neuchâtel terminology is at variance with other terminologies is in using the term “classification” for one particular and well defined concept, and for making a clear distinction between “classification” and “classification version” as explained in previous paragraphs. This and a few other instances of inconsistency are due to the particular focus and purpose of the Neuchâtel Terminology, which calls for quite specific and narrowly, defined concepts.

The nearest in scope to the Neuchâtel terminology is the conceptual data model of the EDI message CLASET, which was developed within the framework of EDI Expert Group 6 (EEG6) of the European Board for EDI Standardisation (EBES) launched by Eurostat. CLASET is designed to exchange tree structures in a standardized way. Although they have been developed for different purposes and with different methodologies the CLASET data model and the Neuchâtel conceptual model have many object types and attributes in common. The two models have been harmonized, where applicable.

1.5 Implementation

Although the 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.

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 currently being implemented in Statistics Sweden and the Swiss Federal Statistical Office with the aim of setting up, developing and managing their national classification database.

Statistics Denmark and Statistics Norway have used the Neuchâtel model to develop a standard relational classification database. This implementation is described in appendix 1.

The implementation of the Neuchâtel terminology in Bridge^{NA} is described in greater detail in appendix 2.

1.6 Layout of the terminology

Section 2 gives an overview of the object types, including a short description. The list is ordered according to an obvious and simple logic. The following pages (section 3) contain 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. A few examples have been added to facilitate understanding. 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 terminology.

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 terminology, e.g. classification, classification version, 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.

A simplified object graph in section 4 gives an overview of the main object types and relationships in the conceptual model.

1.7 References

CLASET. *Electronic Format for Exchange of Classifications*. Eurostat
<http://europa.eu.int/comm/eurostat/ramon/>

CODED. *The Eurostat Concepts and Definitions Database*. Eurostat.
<http://europa.eu.int/comm/eurostat/ramon/>

Hoffman, E. and Chamie, M. *Standard Statistical Classifications: Basic Principles*. Paper presented to the 30th session of the UN Statistical Commission, New York, 1999.
<http://unstats.un.org/unsd/class/family/bestprac/>

ISO 704: 2000. *Terminology Work - Principles and Methods*.

Sundgren, B. *Information Systems Architecture for National and International Statistical Offices Guidelines and Recommendations*. Conference of European Statisticians, Statistical Standards and Studies No 51. United Nations 1999.

Terminology on Statistical Metadata. Prepared by the UN/ECE Work Session on Statistical Metadata (METIS). Conference of European Statisticians, Statistical Standards and Studies No 53. Geneva 2000.

UN Glossary of Classification Terms. Working document. United Nations.
http://unstats.un.org/unsd/class/family/glossary_short

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.

2 Object types

N.B. 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 comprises a number of classifications, which are related from a certain point of view.

Classification: Classification describes the ensemble of one or several consecutive classification versions. It is a "name" which serves as an umbrella for the classification version(s).

Classification version: A classification version is a list of mutually exclusive categories representing the version-specific values of the classification variable. A classification version has a certain normative status and is valid for a given period of time.

Classification variant: A classification variant is based on a classification version. In a variant, the categories of the classification version are split, aggregated or regrouped to provide additions or alternatives to the standard order and structure of the base version.

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 classification version or variant.

Correspondence table: A correspondence table expresses the relationship between different versions or variants of the same classification or between versions or variants of different classifications.

Classification level: A classification structure (classification version or classification variant) is composed of one or several levels. In a hierarchical classification the items of each level but the highest (most aggregated) level are aggregated to the nearest higher level. A linear classification has only one level.

Classification item: A classification item represents a category at a certain level within a classification version or variant.

Item Change: One or more items of a classification version or variant can be subject to a change. Item change describes the change that has occurred.

Case law: A case law is a ruling concerning the classification of certain statistical objects/units, which do not obviously fit into one category and are not clearly defined by explanatory notes.

Classification index entry: An index entry is a word or a short text (e.g. the name of a locality, an economic activity or an occupational title) that describes a type of object/unit or object property to which the classification version or variant applies, together with the code of the corresponding item.

Correspondence item: A correspondence item expresses the relation between an item of the source version and a corresponding item of the target version. Any item of a source version or a target version may be related to several correspondence items.

3 Object types and attributes

3.1 Classification family

A classification family comprises a number of classifications, which are related from a certain point of view. The family may be based, for instance, on a common classification variable (e.g. economic activity) or on a common type of object/unit. 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.

Title: A classification family has a title.

Ex.: "Geographical classifications"; "Product classifications".

Classifications: A classification family refers to a number of classifications.

Ex.: A Product classifications family may comprise e.g. CPA, CPC, HS, CN

See also:

Classification

3.2 Classification

Classification describes the ensemble of one or several consecutive classification versions. In the context of the classification database, there is no structured list of categories directly associated with the classification. It is a "name" which serves as an umbrella for the classification version(s).

Identifier: A classification is identified by a unique identifier, which may typically be an abbreviation of its title.

Ex.: ISIC, ICD.

Title: A classification has a title as provided by the owner.

Ex.: ISIC - International Standard Industrial Classification; ICD - International Classification of Diseases.

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

Context: A classification can be designed in a specific context.

Ex.: ISIC: international classification; NACE: EU classification; NACE-BEL: Belgian classification.

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

Ex.: Enterprises by economic activity, products by origin, persons by age.

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

Ex.: ISCO is used in employment and labour force statistics.

Owners: The statistical office or other authority, which created and maintains the version(s) of the classification. A classification may have several owners.

Ex.: ISIC is owned by UNSD; ISCO-COM is owned by ILO and Eurostat.

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

Ex.: For NACE: "Economic Activity"; "Industry"; "Production".

Family: Classifications may be grouped into classification families. Shows to which family the classification belongs.

Ex.: ISIC and NACE belong to the family "Classifications of economic activity".

Versions: A classification has at least one version (classification version).

Ex.: ISIC: ISIC Rev.1, ISIC Rev.2, ISIC Rev.3; NACE: NACE 70, NACE Rev.1.

Current version: If there are several versions of a classification, one version *may be* assigned as the currently valid version.

Ex.: ISIC Rev. 3; NACE Rev. 1.

Footnotes

See also:

Classification version, Classification family

3.3 Classification version

A classification version is a list of mutually exclusive categories representing the version-specific values of the classification variable. If the version is hierarchical, each level in the hierarchy is a set of mutually exclusive categories. A classification version has a certain normative status and is valid for a given period of time. A new version of a classification 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 the older version. Border changes may be caused by creating or deleting categories, or moving a part of a category to another. The addition of case law, changes in explanatory notes or in the titles do not lead to a new version.

Identifier: A classification version is identified by a unique identifier, which may typically be an abbreviation of its title. It is often distinguished from other versions of the same classification by reference to a revision number or to the year of the version's coming into force.

Ex.: NACE Rev.1, CPA 96.

Title: A classification version has a title as provided by the owner or maintenance unit.

Ex.: Classification of Products by Activity, 1996.

Description: Short general description of the classification version.

Release date: Date on which the classification version was released.

Termination date: Date on which the classification version ceased to be valid.

Current version: Indicates whether or not the classification version is the currently valid version.

Introduction: The introduction provides a detailed description of the classification version, the background for its creation, the classification variable and objects/units classified, classification rules etc.

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

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

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

Ex.: The legal base for NACE Rev.1 is:

1. Council of the European Communities Regulation (EEC) No. 3037/90 of 9 October 1990 (Official Journal of the European Communities No. L 293, 24.10.1990)
2. Commission of the European Communities Regulation (EEC) No. 761/93 of 24 March 1993 (Official Journal of the European Communities No. L 83, 3.4.1993)

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

Title types: A list of the defined types of alternative item titles available for the version. Each title type refers to a list of alternative item titles.

Ex.: Short titles; Medium titles.

Languages available: A classification version 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.

Floating: If a classification version is updated often, it can be marked as a floating classification version. In a floating classification version a validity period should be defined for all items which will allow the display of the item structure and content at different points of time. **Updates possible:** Indicates whether or not updates are allowed within the classification version, i.e. without leading to a new version. Such updates will usually be changes, which add items to the structure and/or revalidates/invalidates classification items but which do not alter the borders within the structure of the classification versions.

Changes from previous version: Summary description of changes which have occurred from the previous classification version to the actual version.

Updates: Summary description of changes which have occurred within the classification version. **Copyright:** Classification versions may have restricted copyrights. Such versions might be excluded from downloading and should be displayed in official publications (e.g. WEB) indicating the copyright owner.

Dissemination allowed: Indicates whether or not the classification version may be published or disseminated (e.g. on the Web).

Classification: A classification version is a version of one specific classification.

Ex.: CPA 96 is a version of CPA.

Predecessor version: The nearest previous classification version to the actual version.

Successor version: The classification version immediately following the actual version.

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

Ex.: NACE Rev.1 is derived from ISIC Rev.3; CPA 93 is derived from the Provisional CPC (1991).

Levels: The structure of a classification version is defined by its levels (classification level).

Items: A classification version 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.

Case laws: A list of all case laws associated with the classification version.

Correspondence tables: A classification version may be linked to other classification versions or classification variants through correspondence tables.

Ex.: Correspondence table between NACE Rev.1 and ISIC Rev.3.

Classification indexes: A classification version can be associated with one or a number of indexes in which index entries are linked to the appropriate classification item.

Footnotes

See also:

Classification, Classification variant, Classification level, Classification item, Case law, Correspondence tables, Classification index,

3.4 Classification variant

A classification variant is based on a classification version. In a variant, the categories of the classification version are split, aggregated or regrouped to provide additions or alternatives (e.g. context-specific additions) to the standard structure of the base version. A classification variant has two or more levels, which are either base version levels (i.e. levels of the classification version on which the variant is based) or the new levels created for the variant. At least one of the levels must be (part of) a base version level, which defines the relation of the variant to the base version. The term variant does not refer only to the partial structure composed of the new variant levels. It refers to the whole new structure composed of both base version levels and new variant levels.

Variants are commonly of three kinds. These have been named **extension variants**, **aggregate variants** or **regrouping variants**. There could exist other types of variants.

Extension variant: An extension variant is a classification variant which extends the base classification version with one or several new levels at the bottom of the base classification version, creating a new lowest level. An extension variant thus adds new lower levels to the base classification version but does not otherwise alter its original structure.

Aggregate variant: An aggregate variant is a classification variant which groups the categories of a linear classification version to create one or several aggregate level(s), thus creating a hierarchy.

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

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

Regrouping variants which violate the structure of the base version: This type of regrouping variant introduces a new level or new levels on top of any but the topmost level of a hierarchical classification version by regrouping categories of the base classification version in a way which violates its original order and structure. This regrouping variant consists of all classification levels of the base version 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 version level as parent level.

In all variants but regrouping variants, which violate the structure of the base version, all levels of the base version are retained and one or more new levels are inserted. In regrouping variants, which violate the structure of the base version, one or more new levels are inserted and only the base version levels below the new variant levels are retained.

Identifier: A classification variant is identified by a unique identifier that may typically refer to (contain) the identifier of the base classification version.

Title: A classification variant has a title

Description: A classification variant may contain a general description.

Release date: Date on which the classification variant was released.

Termination date: Date on which the successor classification variant of the actual classification variant was released.

Current variant: Indicates whether or not the classification variant is the currently valid classification variant.

Introduction: The introduction provides a detailed description of the classification variant, the background for its creation, classification rules etc.

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

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

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

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

Title types: A list of the defined types of alternative item titles available for the variant. Each title type refers to a list of alternative item titles.

Ex.: Short titles; Medium titles.

Languages available: A classification variant can exist in one or several languages. Indicates the languages available.

Floating: If the base classification versions of a classification variant is floating, the variant should also be marked as a floating classification variant. In a floating classification variant the item structure and content at different points of time can be displayed.

Updates possible: Indicates whether or not updates are allowed within the classification variant, i.e. without leading to a new variant. Such updates will usually be changes, which add items to the structure and/or revalidates/invalidates classification items but which do not alter the borders within the structure of the classification variants.

Changes from previous variant: Summary description of changes which have occurred from the previous classification variant to the actual variant.

Updates: Summary description of changes which have occurred within the classification variant.

Copyright: Classification variants may have restricted copyrights. Such variants might be excluded from downloading and should be displayed in official publications (e.g. WEB) indicating the copyright owner.

Dissemination allowed: Indicates whether or not the classification variant may be published or disseminated (e.g. on the Web).

Base classification version: A classification variant is based on one specific classification version.

Predecessor variant: The nearest previous classification variant to the actual classification variant.

Successor variant: The classification variant immediately following the actual classification variant.

Levels: The structure of a classification variant is defined by its levels. The levels are ordered, i.e. each level has a level number within the variant. For the base version levels, this number might not be the same as the number given to the level in the base classification version.

Base version levels: The levels inherited from the classification version on which the variant is based and which bind the variant to the originating version.

Items: A classification variant 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.

Case laws: A list of all case laws associated with the classification variant.

Correspondence tables: A classification variant may be linked to classification versions or other classification variants through correspondence tables. Besides existing correspondences associated with the base version levels of the variant, correspondences can also be built with the variant levels as source and target levels.

Classification indexes: A classification variant can be associated with one or a number of indexes in which index entries are linked to the appropriate item.

Footnotes

See also:

Classification version, Classification level, Classification item, Case law, Correspondence tables, Classification index

3.5 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 classification version or variant. Indexes serve a general information purpose as extended descriptions of the contents of items and are especially used as support tools when classifying statistical objects/units.

Identifier: An index is identified by a name. If there are several indexes for a classification version they will usually be indexes in different languages. In this case the language should be part of the index name.

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

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

Floating: If the classification version of a classification index is a floating, the classification index should also be marked as a floating classification index. In a floating classification index the classification index entries at different points of time can be displayed.

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

Classification version/variant: An index is related to one particular classification version or variant.

Index entries: A list of all classification index entries associated with the classification version or variant.

Footnotes

See also:

Classification version, Classification variant, Classification index entry

3.6 Correspondence table

A correspondence table expresses the relationship between different versions or variants of the same classification or between versions or variants of different classifications. In the first case the correspondence table links two versions or variants of a classification to facilitate comparability over time. In the second instance it expresses the relationship between different classifications. Correspondences are bi-directional. Correspondence tables may also be used for expressing validation rules.

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

Title: A correspondence table has a title as provided by the owner.

Description: The description contains information about the scope and aim of the correspondence 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 version: The classification version or variant from which the correspondence is made.

Target version: The classification version or variant to which the correspondence is directed.

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

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

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

Source complete: Indicates that all source items (on the source level) correspond to one or more target items.

Target complete: Indicates that all target items (on the target level) correspond to one or more source item.

Floating: If the source and/or target classification versions of a correspondence table are floating classification versions, the correspondence table should also be marked as a floating correspondence table. In a floating correspondence table the correspondences at different points of time can be displayed.

Footnotes

See also:

Classification version, Classification variant, Classification item, Correspondence item

3.7 Classification level

A classification structure (classification version or classification variant) is composed of one or several levels. In a hierarchical classification the items of each level but the highest are aggregated to the nearest higher level. A linear classification has only one level.

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.

Ex.: Sections, Sub-sections, Divisions, Groups and Classes in NACE Rev.1.

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

Ex.: NACE sub-sections were introduced to provide more detail at an aggregate level.

Number of items: The number of items (categories) at the level.

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

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

Ex. In NACE Rev.1 the code structure at the Class level is 99.99

Level coverage: Indicates whether the level covers the whole range of the classification version or only part of it. Incomplete coverage means that one or several items at the nearest higher level are not further subdivided.

Ex.: Only 2 of 17 Sections of NACE Rev.1 are subdivided into sub-sections.

Dummy code: Rule for the construction of dummy codes from the codes of the next higher level when an incomplete level is made complete.

Ex.: The NACE Rev.1 subsection dummy codes may be created by doubling the section code letter, i.e. Section code: B, subsection dummy code: BB

Foreign level: A foreign level is a level that has been inherited from another classification version and is owned and maintained by the owners and maintenance unit of that classification version.

Items: An ordered list of the categories (classification items) that constitute the level.

See also:

Classification version, Classification variant, Classification item

3.8 Classification item

A classification item represents a category at a certain level within a classification version or variant. It defines the content and the borders of the category. A statistical object/unit can be classified to one and only one item at each level of a classification version or variant. An item is associated with a code and a title and may include explanatory notes.

Code: A classification item is identified by an alphabetical, numerical or alphanumerical code, which is in line with the code structure of the classification level. The code is unique within the classification version or variant to which the item belongs.

Official title: A classification item has a title as provided by the owner or maintenance unit. The title describes the content of the category.

Alternative titles: An item can be expressed in terms of one or several alternative titles. Each alternative title is associated with a title type.

Ex.: Short titles; Medium titles; Self-explanatory titles in CN; Titles in plural form (e.g. Men, Women) for dissemination purposes; gender related titles.

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 item belongs.

Generated: Indicates whether or not the item has been generated to make the level to which it belongs complete.

Ex.: In NACE Rev.1 one may generate items AA, BB, FF etc. to make the subsection level complete.

Currently valid: If updates are allowed in the classification version or variant, an item may be restricted in its validity, i.e. it may become valid or invalid after the classification version or variant has been released. Indicates whether or not the item is currently valid.

Valid from: Date from which the item became valid. The date must be defined if the item belongs to a floating classification version.

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

Past events: The past events describe a change (or a number of changes) that have created the item e.g. from one or several predecessor items (see Item Change). This allows tracing of the history of an item including changes such as new code, fusion or splitting.

Future events: The future events describe a change (or a number of changes) related to an invalid item. These changes may e.g. have turned the now invalid item into one or several successor items (see Item Change). This allows the possibility to follow successors of the item in the future.

Changes from previous version: Describes the changes, which the item has been subject to from the previous to the actual classification version or variant. **Updates:** Describes the changes, which the item has been subject to during the life time of the actual classification version or variant.

Classification version/variant: The classification version or variant to which the item belongs.

Parent item: The item at the next higher level of the classification version or variant of which the actual item is a sub item.

Ex.: In NACE Rev.1 item 10 is the parent of item 10.1.

Sub items: Each item, which is not at the lowest level of the classification version or variant, might contain one or a number of sub items, i.e. items at the next lower level of the classification version or variant.

Ex.: In NACE Rev.1, the Group level items 10.1, 10.2 and 10.3 are sub items of the Division level item 10.

Linked items: Items of other classification versions or variants with which the item is linked, either as source or target, through correspondence tables.

Ex.: CPA 96 item 22.11.31 is linked with CPC 91 item 32220 and HS 96 item 4901.91

Case laws: Refers to one or more case law rulings related to the classification item.

Index entries: A list of all the classification index entries, which are associated with the item.

Footnotes

See also:

Classification level, Classification version, Classification variant, Case law, Classification index, Classification index entry

3.9 Item Change

One or more items of a classification version or variant can be subject to a change. The item change describes the change that has occurred.

Date of change: Date when the change occurred

Type of change: For a given classification version or variant, a set of change types may be defined and recorded (see more extensive definitions of change types in Appendix 3):

deletion

creation

merger

take-over

breakdown

split off

transfer

removal

code change

name change

Description: Text describing reason for, starting point and result of the change.

Predecessors: Predecessor items describe the source item(s) for the change.

Successors: The successor item(s) are the items resulting from the change

See also:

Classification item

3.10 Case law

A case law is a ruling concerning the classification of certain statistical objects/units, which do not obviously fit into one category and are not clearly defined by explanatory notes. A case law is typically related to one classification version or variant only. It may be associated with one or several classification items.

Description: Detailed description of the case law.

Items: Classification items associated with the case law.

See also:

Classification item, Classification version, Classification variant

3.11 Classification index entry

An 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 the classification version or variant applies, together with the code of the corresponding classification item. Each index entry typically refers to one item of the classification version or variant. Although an index entry may be associated with an item at any level of the classification version or variant, index entries are normally associated with items at the lowest level.

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

Item: The classification item with which the 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.

See also:

Classification item, Classification index

3.12 Correspondence item

A correspondence item expresses the relation between an item of the source version and a corresponding item of the target version. Depending on the relationship type of the correspondence table, there may be several correspondence items for a single source or target item.

Source item: The source item refers to the classification item in the source version.

Target item: The target item refers to the classification item in the target version.

Valid from: Date from which the correspondence item became valid. The date must be defined if the correspondence item belongs to a floating correspondence table.

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

Valid from: This information is usually used for floating correspondence tables. It defines the date from which the correspondence item becomes valid.

Valid to: This information is usually used for floating correspondence tables. It defines the date from which the correspondence item becomes invalid.

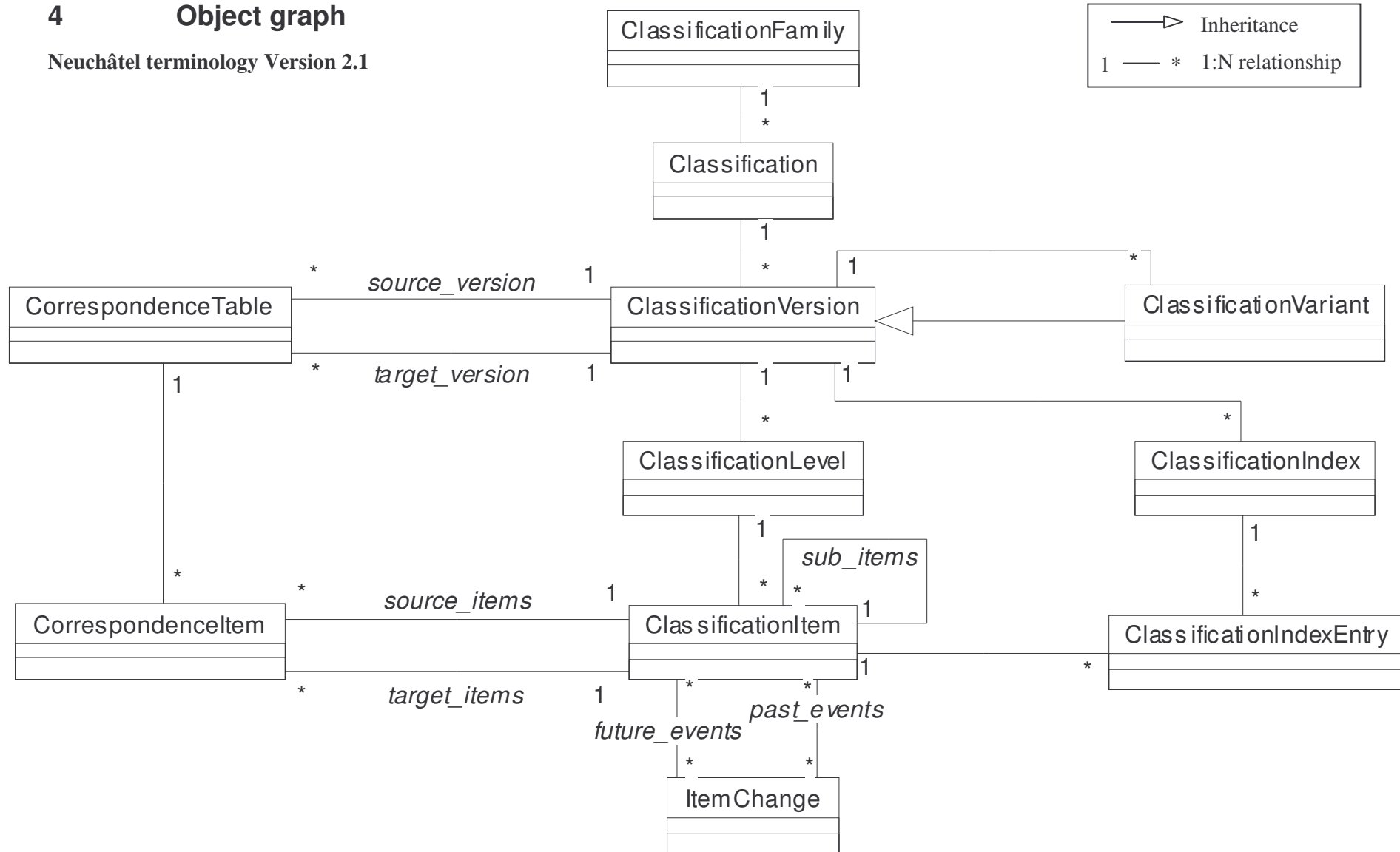
Footnotes

See also:

Classification version, Classification variant, Classification item, Correspondence table

4 Object graph

Neuchâtel terminology Version 2.1



Appendix 1.

Implementation in Statistics Denmark and Statistics Norway.

Statistics Denmark and Statistics Norway decided in the spring of 2001 to start a joint project with the goal to implement the Neuchâtel model in a multilingual standard relational database.

The project implements Version 2.0 of the Neuchâtel specifications with a few attributes being left out.

The model has been implemented using Microsoft Access Database for demonstration and testing purposes and using Oracle for production purposes.

An application in Visual Basic was developed for maintaining the database (creating and updating classifications including variants, indexes and correspondence tables). After user testing in Statistics Norway, production started in July 2002.

Import and export functions have been defined to a preliminary XML format (used as a basis for discussion about a common format within the Neuchâtel group). Other import formats have been established including the option to import from MDB files downloaded from Eurostats Ramon Classification site.

In addition a WEB-based browser has been tested and is now used for the dissemination of selected classifications on the Intranet.

During the implementation, the benefits of having a terminology providing a conceptual framework and defining the basic object types have been evident. The Neuchâtel model has proved to be a solid foundation for the modelling of the database. It is not, and should not be, a complete model, as some attributes may be implemented in different ways according to the environment. As an example, the attribute Publications in this model is implemented as a string (pure text). If a catalogue of publications were available, it might have been implemented as a (set of) pointer(s) into that catalogue.

With the limited time and resources available for the project, the results described above could not possibly have been achieved without the Neuchâtel model as a base.

Appendix 2.

Implementation in the Bridge^{NA} metadata software

Run Software GmbH in Berlin has developed the Bridge^{NA} metadata software, which is a terminology based software build on a Semantic Interface that allows communicating without dependency on specific database software or a specific metadata model. A semantic interface is based on common terminology definitions (i.e. on the conceptual metadata models as defined by statistical experts). The interface provides access mechanisms based on different technical standards as CORBA or COM.

The semantic interface maps the conceptual model to different metadata model implementations. Thus, statistical software becomes independent of specific implementations and models for statistical metadata databases.

ComeIn - a technical solution for implementing standard terminology

The common metadata interface (ComeIn) is an implementation technique for standard terminologies. ComeIn is based on a mapping between technical and conceptual terms and properties. Referring to a standard term (e.g. "Classification") allows accessing metadata for classifications independently on how they are named in the metadata database of the national statistical office. ComeIn is designed in a way that supports terminology extension at any time and is able to describe any number of conceptual terms (logical objects).

Semantic Interface ComeIn

The technical interface specification of ComeIn is public as well as the COM implementation of the interface, which can be downloaded from the WEB site www.run-software.com.

ComeIn supports a number of different platforms by providing a C, C++, COM and a CORBA interface. Providing ComeIn for a special metadata environment means providing the ComeIn DLL that maps the conceptual object types and properties to the specific database implementation.

The ComeIn interface supports multi-lingual approaches as well as version (time) dependent metadata

Classification editor ClassE

ClassE is an enhanced classification Editor developed as one of the tools for the Bridge^{NA} system . It provides a user interface for browsing, editing and importing or exporting classifications, correspondence tables, classification indexes etc. Textual information is multilingual. A number of import/export functions support data exchange between ClassE and Word documents, Excel, relational databases or XML files.

ClassE has been developed by run Software GmbH in cooperation with Statistics Norway, Statistics Sweden, The Swiss Federal Statistical Office and Statistics Denmark. The terminology for classifications and related objects is based on the standard terminology of the Neuchâtel group.

Appendix 3 A TYPOLOGY OF ITEM CHANGES

1. Introduction

Correspondence Tables (3.6) express the relationships between subsequent versions of a classification. These relationships show how the items of the target/successor version relate to the items of the source/predecessor version. 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. The typology also applies for classifying (often minor) changes *within* a particular version, which is then called a *floating version*.

Before presenting the typology, we have to make clear what we understand by “change of a classification item”.

2. Features of Item Change

A classification item is the designation of a category of a classification. It can be said to consist of three components:

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

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 the extensional part, its denotation.

Ideally, naming, coding and definition ought to behave consistently, while the definition is “leading”, in the sense that a change of definition² 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

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

² notice that not all changes in (the wording of) intensional 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 an *index* or *case law*.

A. REAL CHANGE

A.1. DELETION (1 : 0)



An item expires, while its denotation reduces to zero; it does not proceed as (part of the) denotation of one or more other (existing) items.

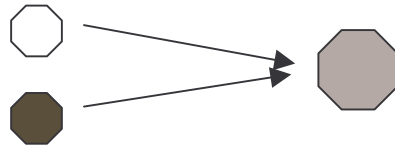
A.2 CREATION (0 : 1)



The mirror image of deletion: an item emerges, while its denotation is not (part of) the denotation of one or more existing items.

A.3 COMBINATION (N:1)

A.3.1 MERGER



Two ore more items expire, while their denotations proceed in one emerging item.

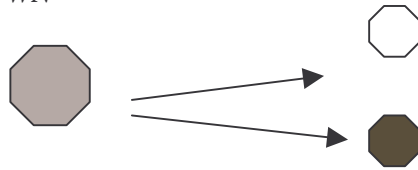
A.3.2 TAKE-OVER (N: 1)



An item expires, while its denotation proceeds as part of the denotation of another item, which continues its existence.

A.4 DECOMPOSITION (1: N)

A.4.1 BREAKDOWN



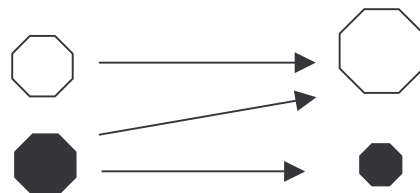
The mirror image of merger: an item expires, while its denotation is distributed over and proceeds in two or more emerging items.

A.4.2 SPLIT OFF



The mirror image of take-over: an item emerges, while its denotation is part of the denotation of another item, which continues its existence.

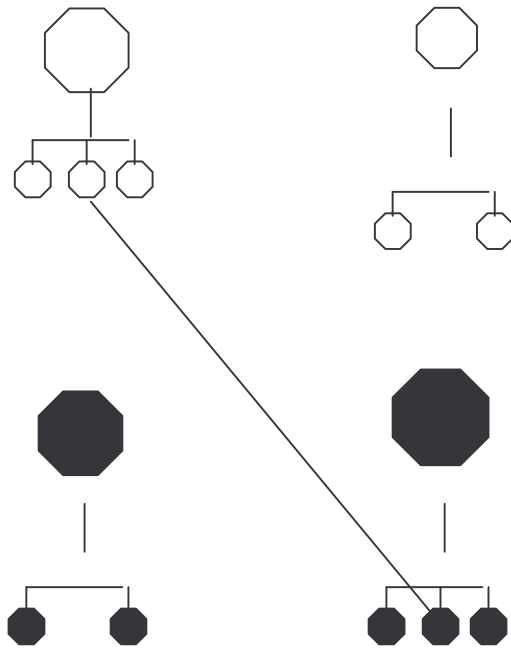
A.5 TRANSFER (M : N)



Part of the denotation of an item moves to another (existing) item

Note:

1. The situation pictured is the most simple of a number of situations, in which more items may be involved in the relationship between the old and new structure.
2. "Transfer" can also be applied at higher levels of a classification. In such cases *the part of the denotation that moves* (to another (existing) item) corresponds to an item at the lower level:



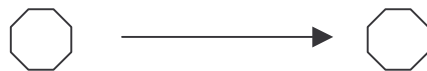
B. VIRTUAL CHANGE

B.1 CODE CHANGE³ (1: 0 ; 0 : 1)



An item expires, while its denotation proceeds as the denotation of an emerging item.

B.2 NAME CHANGE (1 : 1)



The name of an item changes, while its denotation remains the same.

³ Strictly spoken, this heading is not correct, because different codes denote different items.