ESA/STAT/AC.267/9
26 April 2013

UNITED NATIONS
DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS
STATISTICS DIVISION

Expert Group Meeting on International Statistical Classifications
New York, 13-15 May 2013

Neuchâtel Terminology Model: Classification database object types and their attributes Revision 2013

Anne Gro Hustoft, Statistics Norway, and Alice Born, Tim Dunstan and Debra Mair, Statistics Canada

INF.2
12 April 2013

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Work Session on Statistical Metadata
(Geneva 8-10 May 2013)

Topic (i): Metadata standards and models
Topic (ii): Metadata in the statistical business process
Topic (iii): Case studies and tools

Neuchâtel Terminology Model: Classification database object types and their attributes Revision 2013

Invited Paper

Prepared by Anne Gro Hustoft, Statistics Norway, and Alice Born, Tim Dunstan and Debra Mair, Statistics Canada

I. Background

The statistical offices of Denmark, Norway and Sweden were among the participants of the IMIM (Integrated Metainformation Management System) project of the 4th Framework Programme of the EU. The main result of the IMIM project was a software product, BRIDGE, an object-oriented system for metadata management. 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 Run Software-Werkstatt (developers of the BRIDGE software) focussing on the classification database part of BRIDGE. This was the start of the "Neuchâtel group". The aim of the Neuchâtel group was to clarify some basic concepts and to arrive at a common terminology for classifications. The terminology 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 work lists and describes the typical object types of a classification database, and the attributes connected with each object type. 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 development of the model had a practical focus as all of the participating National Statistical Offices (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.

Later, Statistics Netherlands joined the Neuchâtel group, and a new version of the terminology, version 2.1, was released in 2004. In this version "Item change" was introduced as a new object type, and the attribute "floating" was added to the attributes of classification version, classification variant, classification index and correspondence table. Floating means that a validity period should be defined for all classification items. This allows the display of the classification item structure and content at different points of time, for example, in a classification version.

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.

The Neuchâtel terminology was developed by the members of the Neuchâtel group, but the group does not formally exist anymore. Several of the members have retired and others have changed their place of employment. It was always an important premise for the group that the work should be public and available to anyone free of charge.

II. Why revise now?

Many countries have at least partially implemented the model. These include Austria, Belgium, Bulgaria, Canada, Croatia, Czech Republic, Denmark, Estonia, Germany, Greece, Ireland, Norway, Portugal, Slovak Republic, Slovenia, Sweden, Switzerland and the Netherlands. Many of these implementing countries have expressed a desire to see some revisions to the model. There is an opportunity to learn from their experiences what parts of the model need to be revised.

At the 2011 METIS Workshop, participants discussed the need to revise the Neuchâtel Model for classifications. Subsequent to the workshop, it was decided by the METIS Steering Group to contact 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.

There was also impetus from the High-level Group for the Modernization of Statistical Production and Services and the projects that have been spawned by it. For example, the development of the Generic Statistical Information Model (GSIM) draws heavily from the Neuchâtel Terminology Model for the information objects in the GSIM Concepts Group. This exercise further brought to the fore the need to clarify and more completely explain certain parts of the Neuchâtel Model and how they relate to each other.

III. Revision process

A joint working group with members from the UNECE METIS Group for Metadata and the UN Expert Group on International Statistical Classifications has explored the need for revision and is currently developing specific revision proposals. The working group is comprised of members from Australia (Helen Toole, Caley Forrest, Erica Tolcvay), Austria (Norbert Rainer), Canada (Alice Born, Debra Mair and Tim Dunstan), France (Benoit Rouppert), New Zealand (Andrew Hancock), Norway (Anne Gro Hustoft and Jenny Linnerud), Portugal (Isabel Valente), Sweden (Klas Blomqvist), Switzerland (Angelina Dungga Winterleitner), and the United States (Dan Gillman). Members from international organisations

include - Eurostat (Ana Franco), ILO (David Hunter), UNECE (Fiona Willis-Nunez), and UNSD (Ralf Becker).

The first meeting of the working group took place in early December, 2012. At this meeting the scope of the revision was discussed. The working group decided that the revision process would focus on Part 1 of the Neuchâtel Terminology Model, Classification database object types and their attributes. The revision process could consider part 2, the variable model, later, in a second work plan. It was also agreed that the revision should take into consideration the mapping of Neuchâtel terminology into that of GSIM, DDI and SDMX.

It was decided that information on potential revisions would be collected through a questionnaire distributed to all members of this working group and to other countries implementing the Neuchâtel Terminology Model. This amounted to 25 countries and international organisations. The questionnaire was developed by members from Canada, Norway, and Sweden and distributed by the third week of December 2012.

A. Questionnaire results

The survey instrument was an electronic questionnaire comprised of four parts.

Introduction:

- Description of the revision project
- Basic respondent information

Use of standards:

- Current use of the Neuchâtel Model for Classifications, Parts I and II, the Generic Statistical Information Model (GSIM), the Data Documentation Initiative (DDI) and the Standard for Data and Metadata Exchange (SDMX).
- Which specific objects in the Neuchâtel Model for Classifications[1] have been implemented or partially implemented by the responding organization in their classification management system.

Revision

- Which of the object types need to be revised? What is the nature of the revision required? What problems or omissions should be addressed?
- What other issues should be addressed in the revision?

Stewardship

- Where should stewardship of the revised Neuchâtel Model for Classifications reside?
- Should the revised terminology be released under a Creative Commons Attribution license?

The questionnaire (Appendix 1) was sent to all countries and international organisations in the working group as well as countries and organisations participating in UNECE METIS thought to be using Neuchâtel and members of the UN Expert Group on International Statistical Classifications.

[1] Neuchâtel Model for classifications: Part I: Classification database object types and their attributes, Version 2.1 is referred to as Neuchâtel Model for Classifications in the remainder of the document.

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.

Summary of results:

Standards and models	Considering Use	Currently in use	Need for revision
Neuchâtel Model for Classifications	4	11	
Neuchâtel Terminology Model: Part II: Version 1.0	2	3	
Generic Statistical Information Model (GSIM)	10		
Data Documentation Initiative (DDI)	7	5	
Standard for Data and Metadata Exchange (SDMX)	9	8	
Neuchâtel			
Classification family	3	13	2
Classification	2	14	4
Classification version	2	14	6
Classification variant	2	9	8
Classification index	3	10	5
Correspondence table	2	14	4
Classification level	2	14	5
Classification item	2	14	5
Item change	1	7	5
Case law	3	4	8
Classification index entry	3	10	4
Correspondence item	2	14	5

Details about the nature of the revisions desired were also provided by respondents and these were taken into consideration in the analysis of responses and the development of the approach and priorities for revision of the model.

Stewardship	
UN Expert Group on International Statistical Classifications	2
UNECE METIS Group on Metadata	3
Joint UN EG on classifications and UNECE METIS	9
Other	3

Licensing	
Creative Commons	Yes - 11

From the questionnaire results, it was determined that the main priorities for revision should be the objects “Classification version”, “Classification variant”, and “Case law”. However, the working group agreed not to focus solely on these priorities but to also address other concerns raised about the model.

A summary of the input received through the questionnaires was sent to all working group members. In order to generate specific revision proposals for discussion, members were invited to write short papers on issues raised in which they had particular interest or expertise. These papers were used to focus the discussion in subsequent conference calls. Short papers were written on: Variants; Variants, Versions and Time Frames; the Introduction to Classification Versions; the relation of the Neuchâtel Model for Classifications to the Generic Statistical Information Model; and, Case Law.

The summary of the input received was expanded to include proposed solutions for the problems raised and where issues still had to be resolved. This was provided to working group members who then provided written input, commenting further on the specific change requests received and accepting or questioning the proposed solutions. On the basis of this feedback, a list of decisions required was identified and the opinion of the working group was then sought by means of a conference call, to allow for thorough discussion. The resulting revision proposals are reflected in this paper and presentations to the METIS 2013 Work Session in Geneva and to the Expert Group on International Statistical Classifications at its May 2013 meeting in New York.

IV. Specific revisions

This section presents some of the major changes currently under consideration. It should be noted that the revision work has not been finalized and the proposals presented here may be altered in the course of further discussion or as a result of feedback received from the METIS group. The paper focuses on changes affecting the following object types: classification version, classification variant, case law and item change, as well as those related to the structure of a classification database.

A. Distinguishing among: a classification variant, a classification version and a new classification

Comments received led the revision working group to focus on clarifying the definitions of the core concepts: classification version, classification variant and new classification. In the Neuchâtel Model for Classifications, the creation of a new version is described as follows:

> From section 3.3 - 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 old version. Border changes may be caused by creating or deleting categories, or moving 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.

Comments received sought clarification as to the types of changes that could occur without requiring recognition of a new version and, on the other hand, the minimum change required for a new version. Was it necessary that the essential change occur at the lowest level of the classification or could it occur at higher levels only?

A variant is currently defined in the Neuchâtel Model for Classifications as having two characteristics:

1. At least one level of the variant is also present in its base classification version.

> From section 3.4 - 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.
> 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.

2. The variant and base classification version are based on similar grouping criteria.

> From section 1.3.3 - 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.

Comments received suggested this definition of a variant was too restrictive. It failed to acknowledge the relationship to the base version of classifications developed from it, that is, built from its categories, but introducing a new (additional) grouping criterion. It also does not recognize as variants classifications built from different levels of the base version. (These examples will be more fully elaborated below.)

In developing an approach to clarify the different relationships that exist between classifications, several dimensions were considered. Specifically, they took into account: the registration status of the classifications; their purpose; and, the nature of the structural differences between them.

Looking first at registration status, one could argue that a classification is a version of an earlier classification if, because of its introduction, the previous classification was superseded. No particular relationship between the structures of the classifications would be required. They could, for example, have identical structures with updated explanatory notes, class titles and index entries. Such a case was observed when Canada's National Occupational Classification for Statistics (NOC-S) 2001 became superseded when the identically structured NOC-S 2006 was formally recognized as the standard occupational classification. On the other hand, a classification that differed structurally but became valid without superseding the pre-existing version, would be a variant.

Purpose or objective could also be used to distinguish between versions and variants. A series of versions can be connected over time as they serve the same main purpose, the same objective. Variants, however, serve a different, more specific, purpose. For example, whereas the versions of the standard industry classification serve, over time, to order information on industry for a broad range of uses, the variant identifying information and communication technology industries was developed to address a more specific information need.

Finally, the distinction between versions and variants could be made with reference to the structural changes made to the first in order to create the second. Here it is argued that variants are created from their base classifications through the creation of new categories and/or the decomposition or breaking up of existing categories. On the other hand, to create a version there must be change of at least one of the following types: 1) deletion, that is, removing a category; 2) combination, that is, merging categories or having one category take over another; or 3) transfer, that is, moving a category from under one parent to another.

These structural criteria for identifying a version closely reflect the current criterion in the Neuchâtel Model for Classifications for a version. In Neuchâtel, there must be an essential change, that is, a change that alters the borders between categories.

However, Neuchâtel currently recognizes the creation of categories as adequate to determine that a new version has been created. Based on the criteria identified above, creation of new categories would result in a variant, not a version. The definition of version as resulting from structural changes is thus less inclusive than the current definition in Neuchâtel.

To what extent are the suggested structural criteria compatible with Neuchâtel's current discussion of "variant"? To answer this, the four types of variants discussed in Neuchâtel will be considered. It is found that the proposed structural criteria would include as variants three of these types. Specifically, it would include:

> Extension variants: Variants that add new lower levels to the base classification version but do not otherwise change the original structure.
>
> Aggregate variants: Variants that group the categories of a linear classification.
>
> Regrouping variants which do not violate the structure of the base version: Variants that add 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.

However, the structural criteria proposed would not include the fourth type currently recognized, regroupings that violate the structure of the base version. These would, therefore become simply new classifications. This conclusion was not viewed as acceptable by members of the working group who questioned the usefulness of not recognizing these classifications as variants of the classification version from which they had been developed.

Related discussions noted that such regroupings may be introducing new grouping criteria. As will be recalled, if the grouping criteria differ too much from those of the base classification version, the result, according to the criteria for "variant" currently in the Neuchâtel Model, is to be considered a new classification. One example cited as illustration of such a regrouping is the re-aggregation of industry classifications into ICT sector industries and non-ICT sector industries. As noted above, the potential exclusion of such cases was seen as overly restrictive.

The ICT sector example also illustrates the second way in which the current definition of "variant" was seen to be too restrictive. The new classes, ICT and non-ICT do not sit over any one level of such industry classifications as the International Standard Industrial Classification (ISIC) or the North American Industrial Classification System (NAICS), but split certain of their major groups on this new criterion. Again, the exclusion of such classifications from "variant" and considering them to be simply new classifications did not seem useful.

The working group concluded that a less constrained definition of "variant" would be preferable.

These deliberations resulted in the following proposed new definitions.

> 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 **from a particular date for a period that may or may not be specified. A new version is created when a classification version is superseded by the introduction of a new 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 titles do not lead to a new version.

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

Floating classifications are considered to constitute one version.

Classification variant
A variant is built from the classification items of a base classification version. These classification items need not all be found at the same level in the base classification version. To these classification items, one or more new levels may be added. This can include extending the base classification version with one or several new levels at the bottom of the base classification version, creating a new lowest level.

It should be noted that variants are typically developed to serve a specific purpose.

B. **Further changes regarding the object type: Classification version**

The working group is considering changes to several attributes associated with the object type Classification version (see Section 3.3 of the Neuchâtel Model for Classifications.) These proposals are discussed below, organized by attribute.

Description
This attribute would be deleted. It was considered redundant to the attribute, Introduction.

Termination date
In recognition that when classification versions cease to be valid they are typically superseded by a successor version, the wording here would be changed to:

> Date on which the classification version was superseded by a successor version or otherwise ceased to be valid.

A working group member noted, for example, that there can be "deceased" classifications that, for whatever reason, are no longer being maintained by the NSO but could still be used by stakeholders.

Introduction
Currently, the Neuchâtel Model for Classifications sets out the desired content of a classification version's Introduction as follows:

> **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.

This may be the only description in the Neuchâtel Model for Classifications that ends with "etc." Further attention to this attribute in the revised Model is desirable as it provides an opportunity to focus on content-related metadata.

It could be argued that the Introduction of a classification version should provide adequate information to support decision-making regarding the classification's fitness for use in coding a particular data set, or regarding the fitness of use of coded data for particular analytic purposes. It should also provide guidance to survey developers regarding the questions that should be included in their surveys in order to support coding of the data received to that classification version.

Many types of content can be imagined as possibly useful in an introduction. What is actually useful in any specific case may vary. For that reason, an appendix will be added to the Neuchâtel Model for Classifications containing a checklist of possible topics to include in an introduction. Users of the model will be asked to consider the relevance to a specific classification version of each of these topics and to include in the introduction all those that are relevant. Some possible topics to be listed in this Introduction checklist are set out in Appendix 2 of this paper.

Floating classifications
Floating classifications are those whose items and structure can change over time without defining a new version. Instead of versioning the classification over all, each item and its content have attached to them a validity period indicating what classification items and what content is currently valid or was valid on any particular past date. Working group discussions of floating classifications indicated the need for clarification and expansion in the description of this attribute. The nature of the required revision has still to be determined. One point to be resolved is the desired scope of this term. Is a floating classification only 1) one that can undergo changes, including essential changes, at any time without requiring a new version, or does it also include, 2) classifications that can be updated with non-essential changes within a version.

Updates possible
This attribute would be revised as follows in order to more fully discuss the issues that should be addressed here.

> 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. Can classification items be added to the structure? Can they be revalidated or invalidated? Such changes are more likely to be permissible in floating classifications. Also indicate whether changes to such things as item names and explanatory notes that do not involve structural changes to the classification are permissible within a version.**

Items
This attribute would be deleted as links to the classification items will be available from the object type, Levels.

Case law

This attribute would be deleted as case law information is better accessed from the classification item to which it is related.

Levels, Correspondence tables, and Classification indexes
The descriptions of these attributes would be expanded to add the explanation that what could be entered here are links to the relevant levels, correspondence tables and classification indexes.

Variants
This attribute would be added.

> Variants: Identify any variants associated with that classification version.

C. **Further changes regarding the object type: Classification variant**

As was the case for Classification version, it is proposed that the Description, Items and Case law attributes would be deleted. The attribute, Owner, would be added as the owner of a variant may not always be the same as the owner of the classification version.

The three kinds of variants, extension variant, aggregate variant and regrouping variant would be kept; however, for compatibility with the revised definition of variant the description of the subtype, Regrouping variants which violate the structure of the base version, would be rewritten as follows:

> This type of regrouping variant introduces a new level or levels into a hierarchical classification version anywhere but above its topmost level by regrouping categories of the base classification version in a way which violates its original order and structure. The categories of the base classification version that are regrouped by the lowest of the new levels may be from one or more levels of the base classification version. This regrouping variant consists of all classification levels of the base version below the categories of the lowest new variant level plus the new variant level(s). In such a regrouping variant, a new variant level cannot have a base version level as parent level.

It would also be noted that a particular variant could be built through a combination of extension and regrouping; it need not be of one kind only.

Changes to several attributes associated with Classification variants are also proposed. These are presented below, organized by attribute.

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

Updates possible
This attribute would be revised as follows.

> Updates possible: Indicates whether or not updates are allowed within the classification variant, i.e. without leading to a new **version of the variant. Indicate here what structural changes, if any, are permissible.**

Base classification version

This attribute description would be revised to acknowledge that a variant, though based on one specific classification version, could be applicable to subsequent versions of that classification.

Levels, Correspondence tables, and Classification indexes
The descriptions of these attributes would be expanded to add the explanation that what could be entered here are links to the relevant levels, correspondence tables and classification indexes. It will also be noted that links may be required for multiple Levels objects, if the variant is applicable to more than one base version.

D. **Case law**

In the Neuchâtel Model for Classifications, case law refers to 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.

Comments received noted that "case law" might not be the most appropriate terminology, questioned the usefulness of this object type and suggested the addition of an attribute identifying the classification versions or variants to which particular rulings applied.

The proposal currently being considered is to delete this object type and to make case law an attribute of classification item, since case law information is better accessed from the classification item to which it is related. It was also noted that case law rulings do not create new versions, since rulings do not change the structure of the classification nor any borderlines between classification items. Nor do they create variants, as rulings should be applied by all users and for all uses of the classification.

Case law is part of classification maintenance work, which has as its general goal making sure that the descriptions of the classification (version or variant) are updated to the new or changed reality of the classification subject. For example, a new emerged activity or product may need to be considered. Such updates of the index items normally take place without the need for a ruling. However, if the new phenomena do not easily fit into one of the classification categories or if the interpretation of which is the correct classification category differs, a case law ruling might be appropriate.

Case law decisions will be taken by the custodian of the classification. Such decisions need to be documented (particularly the rationale for the decision taken) and the rulings should be available for all who have to work with that classification. They should also be available to data users so that they may know the coverage of the classification categories. Best practice for case law rulings would include an opportunity for concerned parties to contribute proposals and for some discussion to take place before a ruling is set.

Case law rulings should be taken into consideration as inputs to the next revision or update of the classification as they suggest a deficiency in the current classification version.

E. **Item change**

It is proposed that this object type would be deleted. Though the identification of change types was seen as useful, they are fully explained in the Neuchâtel Model for Classifications, Appendix 3, A Typology of item changes. The most appropriate location in which to record this information in the database was determined to be within the Classification item object type. This object type already has the attributes: Changes from previous version; and, Updates in which to record, respectively, changes to which the item has been subject since the previous version and changes to which it has been subject during the lifetime of the present version. The descriptions of these attributes will be expanded to include the instruction to use the typology presented in Appendix 3 when describing these changes. The typology itself will be expanded to include: Change to explanatory note. This addition is required as currently the typology only covers structural changes and makes no reference to changes in the title or descriptive/definitional information associated with a classification item.

F. **Structure of a classification database**

The Neuchâtel Model for Classifications sets out a database structure which is conveyed through the attributes of the object types and, more concisely, through the Object graph. In this structure, the object type Classification index entry is linked to the objects Classification item and Classification index. This is accomplished through including in both Classification item and Classification index the attribute Index entries which contains a list of all the classification index entries associated with that Classification item or index. This results in all index entries appearing three times on the database. As the number of index entries can be very large and the potential for error in entering them multiple times and in maintaining a database that contains such duplication would be significant, it is proposed that such duplication would be eliminated.

In the revision, the attribute Index entries would be dropped from the objects Classification index and Classification item. These additional attributes would be added to the object Classification index entry:

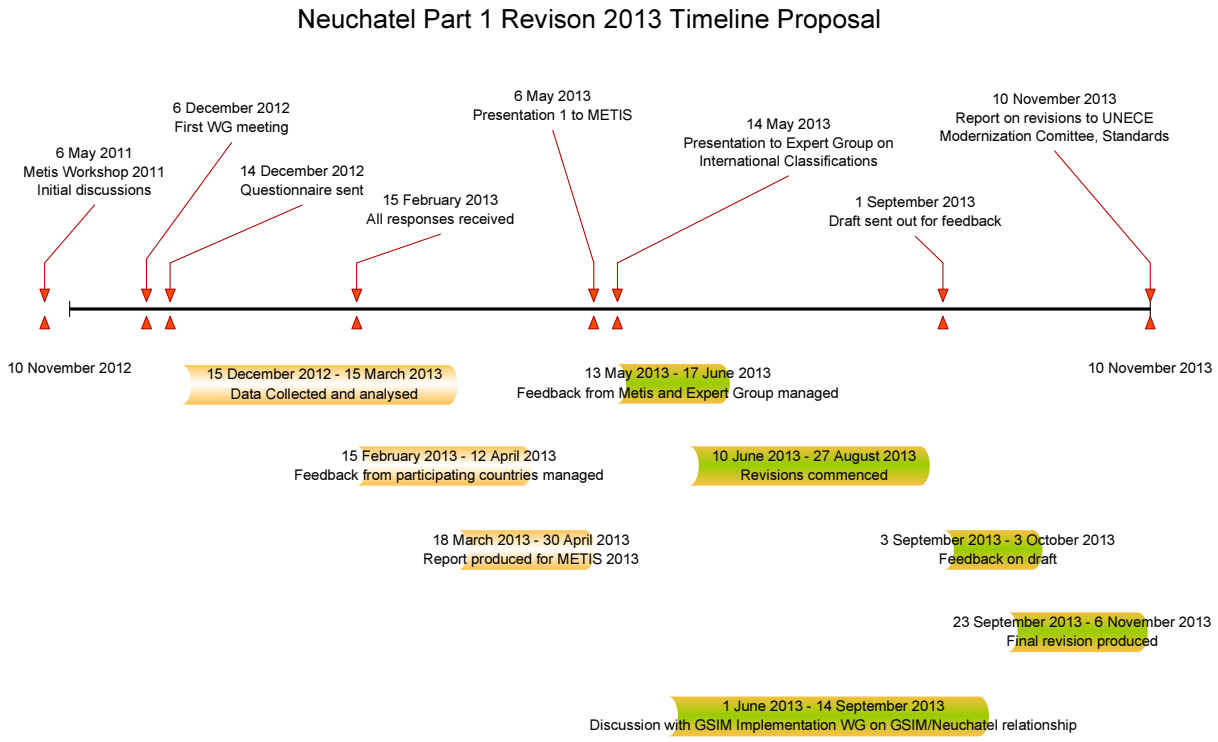
> **Versions/variants: Identify the Classification versions and Classification variants to which the index entry is associated.**
>
> **Codes: For each Classification version or variant to which the index entry is associated, enter the code of the classification item in that version or variant with which the index entry is associated.**

G. **Other changes**

Comments received in the questionnaire responses indicated that users wanted clearer explanations, more examples and better connections to the Generic Statistical Information Model (GSIM). In response to these comments, the document will be generally reviewed to increase clarity and a new appendix will be added in which an example is fully presented. It is currently proposed that another new appendix will contain a mapping of the Neuchâtel Model for Classifications to GSIM which would be based on the mapping currently contained in the Generic Statistical Information Model (GSIM): Specification (Appendix 3).

V. Next steps

The working group is continuing with identification of revisions to the model. Below is a proposed roadmap for the revision.



Before proceeding with any revision to the Neuchâtel Terminology Model Part 1, a number of questions have been raised by the working group members, and need to be addressed.

- A key question for this project is the potential relationship between this work and the ongoing work on the enhancement of GSIM. Should this project aim to produce a new version of the Neuchâtel Model for Classifications or might an expanded GSIM take in the information objects of this Model? We are proposing that the working group work with the GSIM Implementation WG on a resolution.

- Should mappings to other metadata exchange protocols (e.g., SDMX, DDI) be included as part of the model?

- Based on the decisions from point one, is there a need to address the issue of governance and stewardship of the Neuchâtel model. Currently, there is no formal owner of the

model, and this is the first attempt to revise it. While members of the working group supported “joint” stewardship between UNECE METIS – UN EG on classifications, will this be feasible?

Once these issues are resolved, the working group is proposing to finalize the revisions by the end of the year, and determine the appropriate way to publish the revised model. However, there was general agreement at the beginning of the project that the Neuchâtel model remains a useful and robust reference model for specialists who have to manage statistical classifications in NSIs and international organizations.

Appendix 1

December 2012

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

METIS (Work Group on Statistical Metadata)

Questionnaire to identify issues for the revision of the Neuchâtel Terminology Model – Classification database object types and their attributes, Version 2.1

I. INTRODUCTION

The issue of reviewing and potentially revising the Neuchâtel Terminology Model – Classification database object types and their attributes, Version 2.1 has been raised in such international fora as METIS by some of the National Statistical Offices that are using the Model. However, currently, there is no steward for this Model. Because of this, the United Nations Statistical Division (UNSD) Expert Group on International Statistical Classifications and the United Nations Economic Commission for Europe (UNECE) Statistical Division have taken on this revision through a joint working group.

This questionnaire represents the first step in this review and revision process. The questionnaire aims to solicit proposals, descriptions of problems and general comments on the current Model that National Statistical Offices and other users wish to have considered in this revision process. These responses will serve as the basis for determining the scope and content of this revision, and a report to the 2013 METIS Work Session. Your input is therefore extremely important.

Your response is requested by **January 17, 2012** but will be accepted until February 15, 2012.

Your response should be sent to Statistics Canada at:
standards-normes@statcan.gc.ca

Respondent information:	
Country:	
Office:	
Name of contact person:	
Contact info (e-mail):	

1. USE OF STANDARDS

1.1 Please indicate your organization's use of the following standards for statistical classifications:

	No current use	**Considering use**	**Currently in use**	**Use to output data in this format**
Neuchâtel Terminology Model: Part I: Classification database object types and their attributes, Version 2.1	□	□	□	□
Neuchâtel Terminology Model: Part II: Variables and related concepts object types and their attributes, Version 1.0	□	□	□	□
Generic Statistical Information model (GSIM)	□	□	□	□
Data Documentation Initiative (DDI)	□	□	□	□
Standard for Data and Metadat Exchange (SDMX)	□	□	□	□

1.2 Have you developed any documents related to the mapping of the Neuchâtel Terminology Model for Classifications to GSIM, DDI or SDMX? If so, please attach any such materials.

1.3 Which of the objects and their attributes identified in Neuchâtel Terminology Model for Classifications are you using or in the process of using?

Object	No current use	Considering use	Currently in use
Classification family	□	□	□
Classification	□	□	□
Classification version	□	□	□
Classification variant	□	□	□
Classification index	□	□	□
Correspondence table	□	□	□
Classification level	□	□	□
Classification item	□	□	□
Item change	□	□	□
Case law	□	□	□
Classification index entry	□	□	□
Correspondence item	□	□	□

2. REVISION

2.1 Regardless of whether you have implemented the Neuchâtel Terminology Model for Classifications, which of the objects and their attributes in this Model do you think need revision?

Object	Requires revision
Classification family	□
Classification	□
Classification version	□
Classification variant	□
Classification index	□
Correspondence table	□
Classification level	□
Classification item	□
Item change	□
Case law	□
Classification index entry	□
Correspondence item	□

2.2 For each object you checked as requiring revision, please elaborate on the nature of the revision required. What problem or omission should be addressed in the revision?

2.3 Should anything be added to the list of objects? (Please include consideration of current attributes.) If any, why should they be added?

2.4 Are there any other issues not mentioned above that you would like to see addressed in the revision? If so, please elaborate.

3. STEWARDSHIP

3.1 After this revision, where should the stewardship of the Neuchâtel Terminology Model: Classification database object types and their attributes V. 1.2 reside? The steward(s) would take on the responsibility for the maintenance and revision of the model.

□ UNSD Expert Group on International Statistical Classifications

□ UNECE Statistical Division

□ Joint ownership by UNSD Expert Group on Statistical Classifications and UNECE Statistical Division

□ Other, specify

3.1 Do you support release of the revised terminology under a/the creative commons license (http://creativecommons.org/licenses/by/3.0/)?

Appendix 2

Checklist of possible content for the Introduction to a Classification version

Detailed description of the classification version

Describe the structure of the classification, including the number of levels, their names, the number of classes at each level, the structure of the codes and the relationship between the codes used at the different levels.

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

Background for the creation of this classification version

Identify the previous classification version. Reasons for updating could be discussed here. The revision process could be described, including any consultations that occurred.

Relationship of this classification to related classifications

Identify other classifications that classify related subject matter and discuss how the classifications 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 classifications applicable to the same subject matter

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 classification version to relevant international standard classifications

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

Summarize the changes from the previous version. The information provided here would not be as specific as that in the concordance table but would summarize, for example, the number of new classes at each level, the number of classes at each level that were collapsed, the extent to which new index items have been added, the extent to which definitions or class labels have been revised, and identification of any particular sections of the previous classification version that were more extensively revised.

Classification criteria

Identify the criteria on which units have been grouped together in this classification version.

If different classification criteria are used, or given primacy, at different levels of the classification version, 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.

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

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

Include here discussion of the units 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 variable

Identify the underlying variable that is measured/described by this classification. Define this variable and provide any relevant clarifications regarding the scope of the conceptual coverage of the variable 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

Discuss here the information required about a particular unit 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 classification in coding

Describe how best to use the 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.

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

Classification rules

Discuss here any specific rules related to applying the 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.

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

Discuss any rules or principles that have been developed to facilitate the implementation of specific classes. 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

Present the structure of any variants known at the time the classification version is published. Show how they relate to the classification version and discuss the appropriate contexts in which each variant is to be used.

Appendix 3

Mapping the Neuchatel Terminology Model Part 1 to the Classification Model in GSIM V.01

The first version of the Generic Statistical Information Model (GSIM) was released in December 2012. The release of this model was the culmination of a year-long international effort in which representatives from many statistical offices around the world participated in a series of face-to-face and virtual meetings. The output is a set of documents describing the model at several levels of detail and some of the criteria used to develop it.

The model is conceptual, intended to provide a framework through which statistical offices can describe their designs, processes, artefacts, and data, i.e., statistical objects. This includes statistical classifications, categories sets, and code sets – the subject of the Neuchâtel Classification Model (NCM). So, one obvious question is whether the GSIM and NCM describe classifications, category sets, and code sets in a compatible way.

One of the major drivers behind the development of GSIM is the idea of standardizing statistical production. It is not important to lay out all that this means here, but it should be clear this means the standards, models, and other specifications used to describe the work of statistical offices need to be mapped to show, within their scopes, the same kinds of statistical objects are described in translatable ways. If descriptions are not translatable, then there are fundamental incompatibilities among them. This might mean the entire metadata and modeling question must be revisited, this time on the models (e.g., GSIM and NCM) themselves rather than the statistical objects the models describe.

Therefore, the utility of GSIM lies in part on showing it is translatable with other existing statistical metadata standards, models, and specifications. Moreover, not-yet-developed specifications will need to be mapped similarly in the future. In this section, we show that NCM is translatable with GSIM.

Table 1. Mapping between Neuchâtel Terminology for Classifications and GSIM

Neuchâtel terminology	**GSIM**	**Example**	**Comment**
Classification family	*Classification Family*	Activity (Industry) classifications, Educational classifications	Group of Classifications
Classification	*Classification*	NACE, ISIC, ISCO, ANZIC06, NAICS	Group of Classification Schemes
Classification version	*Classification Version*	NACE rev 2, ISIC rev 4, ISCO 08, ANZIC 06, NAICS 2012	
Classification variant	*Classification Variant*	High-level SNA/ISIC aggregation A*10/11 grouping	
Classification level	*Level*	Section, division, group and class in ISIC rev 4	
Classification item	*Classification Item*	0111 - Growing of cereals (except rice), leguminous crops and oil seeds	
Correspondence table	*Correspondence Table*	ISIC rev 4 – NAICS 2012	
Classification index	-		List of aliases
Classification index entry	-		Aliases
Item change	-		
Case law	-		
Classification index entry	Alias on *Node*		
Correspondence item	*Map*	0111 in ISIC rev 4 - 111110 NAICS 2012	
Classification item - code	Attribute on *Classification Item*	0111 (in ISIC)	Not an information object itself in GSIM
Classification item - title	Attribute on *Classification Item*	Growing of cereals (except rice), leguminous crops and oil seeds	Not an information object itself in GSIM
Explanatory notes	Attribute on *Classification Item*	"This class includes: - growing of temporary and permanent crops - cereal grains: rice, hard and soft wheat, rye, barley, oats, maiz corn (except sweet corn) etc.	

		- growing of potatoes, yams, sweet potatoes or cassava - growing of sugar beet, sugar cane or grain sorghum - growing of tobacco, including its preliminary processin harvesting and drying of tobacco leaves - growing of oilseeds or oleaginous fruit and nuts: peanuts, soy colza etc. - production of sugar beet seeds and forage plant seeds (includir grasses) - growing of hop cones, roots and tubers with a high starch or inul content - growing of cotton or other vegetal textile materials - retting of plants bearing vegetable fibres (jute, flax, coir) - growing of rubber trees, harvesting of latex - growing of leguminous vegetables such as field peas and bea growing of plants used chiefly in pharmacy or for insecticida fungicidal or similar purposes - growing of crops n.e.c. *This class excludes:* - *growing of melons, see 0112* - *growing of sweet corn, see 0112* - *growing of other vegetables, see 0112* - *growing of flowers, see 0112* - *production of flower and vegetable seeds, see 0112* - *growing of horticultural specialties, see 0112* - *growing of olives, see 0113* - *growing of beverage crops, see 0113* - *growing of spice crops, see 0113* - *growing of edible nuts, see 0113* - *gathering of forest products and other wild growing material (cork, resins, balsam etc.), see 0200"*	