The Copenhagen Mapping

IMPLEMENTING THE GSIM STATISTICAL CLASSIFICATIONS MODEL WITH DDI LIFECYCLE

May 2014 – Version 0.9

The Copenhagen Mapping describes how the GSIM terminology for statistical classifications maps to DDI 3.2, and offers a set of controlled vocabularies to be used by DDI implementers who wish to describe classifications using the standard.

TABLE OF CONTENTS

The Copenhagen Mapping
Implementing the GSIM Statistical Classifications Model with DDI Lifecycle

1 BACKGROUND

1.1 GSIM STATISTICAL CLASSIFICATIONS MODEL

The Generic Statistical Information Model (GSIM) is the first internationally endorsed reference framework for statistical information. This overarching conceptual framework will play an important part in modernizing, streamlining and aligning the standards and production associated with official statistics at both national and international levels.

The GSIM Statistical Classifications Model is based upon the Neuchâtel terminology model for classification database object types and their attributes v2.1. It was developed by a group of 19 members from 13 different national and international organizations in an endeavor to arrive at a common language and a common perception of the structure of statistical classifications and the links between them. The GSIM Statistical Classifications Model is both a terminology and a conceptual model. It defines the key concepts that are relevant to structuring Statistical Classification metadata and provides the conceptual framework for the development of a Statistical Classification management system [UNECE].

1.2 DDI

The Data Documentation Initiative (DDI) is an effort to create an international standard for describing data from the social, behavioral, and economic sciences. Expressed in XML, the DDI metadata specification supports the entire research data life cycle. DDI metadata accompanies and enables data conceptualization, collection, processing, distribution, discovery, analysis, repurposing, and archiving [DDI Alliance].

2 GAP ANALYSIS

In December 2013, contributors to this paper met in Copenhagen, Denmark and performed a gap analysis to determine whether the DDI Lifecycle standard could be used to implement the GSIM Statistical Classifications Model. Although DDI 3.1 does not provide the necessary elements to fully implement the model, the newly released DDI 3.2 standard is sufficient.

This whitepaper describes how the GSIM Statistical Classifications Model maps to DDI 3.2, and offers a set of controlled vocabularies to be used by DDI 3.2 implementers who wish to describe classifications using the standard.

3 DEFINITIONS FROM GSIM AND DDI

3.1 GSIM TERMINOLOGY
The following definitions are repeated directly from the GSIM Statistical Classifications Model [UNECE].

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

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

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

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

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

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

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

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

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

3.2 RELEVANT DDI 3.2 ELEMENTS

The following DDI 3.2 elements will be used to map the GSIM classifications terminology to the DDI standard.

3.2.1 CodeList

A CodeList contains a list of values associated with categories. The code list may be flat or hierarchical.

3.2.2 CodeListGroup

A CodeListGroup represents a group of related CodeLists. The relationship among code lists in a group can be specified using a controlled vocabulary and a defining concept.

3.2.3 Concept

A DDI Concept describes a concept according to ISO/IEC 11179.

3.2.4 Level

A Level a single level of a hierarchical code list.

3.2.5 Category

A DDI Category is defined using the OECD Glossary of Statistical Terms: Generic term for items at any level within a classification, typically tabulation categories, sections, subsections, divisions, subdivisions, groups, subgroups, classes and subclasses.

3.2.6 CategoryGroup

A CategoryGroup represents a group of related Categories. The relationship among categories in a group can be specified using a controlled vocabulary and a defining concept.

3.2.7 UserAttributePair

A UserAttributePair can be used to annotate a DDI item with user-specified information.

3.2.8 CodeValue

DDI 3.2 allows using formally-specified controlled vocabularies throughout the standard. For the Copenhagen Mapping, these controlled vocabularies are particularly useful for defining types of item groups and for defining extended information stored in UserAttributePair elements.

4 MAPPING GSIM CLASSIFICATIONS TERMINOLOGY TO DDI 3.2

4.1 HIGH LEVEL MAPPING

This section shows how the high level GSIM object types map to DDI elements. For details on how the attributes of each object type map to DDI, see the sections below.

4.1.1 Classification Family: CodeListGroup and Concept

Each Classification Series that belongs to the Classification Family is represented by a DDI CodeListGroup. Each of these CodeListGroups is a member of higher-level CodeListGroup, which represents the Classification Family. The Classification Family's CodeListGroup specifies a Concept that defines the common concept.

4.1.2 Classification Series: CodeListGroup

A Classification Series is represented by a CodeListGroup that contains a CodeList for each Statistical Classification in the series. The CodeListGroup's TypeOfGroup element uses a term from the controlled vocabulary to indicate that it represents a Classification Series.

4.1.3 Statistical Classification: CodeList

A Statistical Classification is represented by a DDI CodeList. The classification is distinguished from regular DDI CodeLists by its membership in a CodeListGroup of the appropriate type, as described above.

4.1.4 Level: Level

A GSIM Level is represented by a DDI Level element within a CodeList.

4.1.5 Correspondence Table: CategoryGroup

A Correspondent Table is represented by a DDI CategoryGroup, which specifies its TypeOfGroup using the provided controlled vocabulary. The CategoryGroup contains one or more nested CategoryGroups, each of which maps equivalent categories to each other.

4.1.6 Classification Index

No direct mapping for a classification index is required. Instead, the functionality can be provided by appropriate tooling, which uses information from the DDI Categories to construct a dynamic classification index.

4.1.7 Classification Item: Code and Category

A Classification Item maps directly to a DDI Code, which consists of a Value and a Category, and can exist within a hierarchical list of items.

4.1.8 Map: CategoryGroup

A Map is represented by a DDI CategoryGroup, which specifies its TypeOfGroup using the provided controlled vocabulary. Each Category Mapping group is nested inside a Correspondence Table group.

4.1.9 Classification Index Entry: Category and UserAttributePair

A Classification Index Entry is represented by a UserAttributePair on a Category, which provides the Coding Instructions that define the classification index entry.

4.2   ATTRIBUTE-LEVEL MAPPING

4.2.1   Classification Family

DDI Item Mapping	
DDI 3.2 Item Type	CodeListGroup
Item Specialization	CodeListGroup -> TypeOfCodeListGroup = cm:ClassificationFamily

GSIM Property	DDI 3.2 Mapping
Identifier	CodeListGroup -> UserId
Name	CodeListGroup -> Name
Classification Series	CodeListGroup -> CodeListGroupReference
Common Concept	CodeListGroup -> ConceptReference

4.2.2   Classification Series

DDI Item Mapping	
DDI 3.2 Item Type	CodeListGroup
Item Specialization	CodeListGroup -> TypeOfCodeListGroup = cm:ClassificationSeries

GSIM Property	DDI 3.2 Mapping
Identifier	CodeListGroup -> UserId
Name	CodeListGroup -> Name
Description	CodeListGroup -> Description
Context	CodeListGroup -> UserAttributePair key = cm:Context
Objects/units classified	CodeListGroup -> UserAttributePair key = cm:UnitsClassified
Subject areas	CodeListGroup -> Subject
Owners	CodeListGroup -> UserAttributePair key = cm:Owner
Keywords	CodeListGroup -> Keyword
Classification Family	(computed relationship from parent Classification Family)
Statistical Classification	CodeListGroup -> CodeListReference

4.2.3   Statistical Classification

DDI Item Mapping	
DDI 3.2 Item Type	CodeList
Item Specialization	CodeList -> UserAttributePair key = extension:type value=cm:StatisticalClassification

GSIM Property	DDI 3.2 Mapping
Identifier	CodeList -> UserId
Name	CodeList -> Name
Introduction	CodeList -> Description
Release date	CodeList -> UserAttributePair key = cm:ReleaseDate
Termination date	CodeList -> UserAttributePair key = cm:TerminationDate
Current	CodeList -> UserAttributePair key = cm:IsCurrent
Maintenance unit	CodeList -> UserAttributePair key = cm:MaintenanceUnit
Contact persons	CodeList -> UserAttributePair key = cm:ContactPerson
Legal base	CodeList -> UserAttributePair key = cm:LegalBase
Publications	OtherMaterials
Name types	CodeList -> CodeListName
Languages available	(computed)
Copyright	CodeList -> UserAttributePair key = cm:Copyright
Dissemination allowed	CodeList -> UserAttributePair key = cm:IsDisseminationAllowed
Classification Series	(computed relationship from parent Classification Series)
Levels	CodeList -> Level
Items	CodeList -> Code
Correspondence Tables	(computed relationship from Correspondence Tables)
Classification Indexes	(computed)
Version	CodeList -> UserAttributePair key = cm:IsVersion
Update	CodeList -> UserAttributePair key = cm:IsUpdate
Floating	CodeList -> UserAttributePair key = cm:IsFloating
Predecessor	CodeList -> UserAttributePair key = cm:Predecessor
Successor	CodeList -> UserAttributePair key = cm:Successor
Derived from	CodeList -> UserAttributePair key = cm:DerivedFrom
Changes from previous version or update	(computed from VersionRationale)
Updates possible	CodeList -> UserAttributePair key = cm:UpdatesPossible
Updates	(computed from VersionRationale)
Variants available	(computed)
Variant	(computed from DerivedFrom)
Changes from base Statistical Classification	(computed)
Purpose of variant	CodeList -> UserAttributePair key = cm:VariantPurpose

4.2.4   Level

DDI Item Mapping	
DDI 3.2 Item Type	LevelType
Item Specialization	none

GSIM Property	DDI 3.2 Mapping
Identifier	CodeList -> UserAttributePair key = cm:LevelIdentifier:<n> where <n> is the integer LevelNumber
Level number	LevelType -> LevelNumber
Level name	LevelType -> LevelName
Description	LevelType -> Description
Number of Classification Items	(computable)
Code type	(computable)
Code structure	(computable)
Dummy code	(computable)
Items	(computable)

4.2.5   Correspondence Table

DDI Item Mapping	
DDI 3.2 Item Type	CategoryGroup
Item Specialization	CategoryGroup -> TypeOfCategoryGroup = cm:CorrespondenceTable

GSIM Property	DDI 3.2 Mapping
Identifier	CategoryGroup -> UserId
Name	CategoryGroup -> Name
Description	CategoryGroup -> Description
Owners	CategoryGroup -> UserAttributePair key = cm:Owner
Maintenance unit	CategoryGroup -> UserAttributePair key = cm:MaintenanceUnit
Contact persons	CategoryGroup -> UserAttributePair key = cm:ContactPerson
Publications	OtherMaterials
Source	CategoryGroup -> UserAttributePair key = cm:SourceClassification datatype = DDI 3.2 urn
Target	CategoryGroup -> UserAttributePair key = cm:TargetClassification datatype = DDI 3.2 urn
Source level	CategoryGroup -> UserAttributePair key = cm:SourceLevel
Target level	CategoryGroup -> UserAttributePair key = cm:TargetLevel
Relationship type	(computed from Map)
Floating	CategoryGroup -> versionDate and (computed via referenced Statistical Classifications)
Maps	CategoryGroup -> CategoryGroupReference

4.2.6   Classification Index

No direct mapping for a classification index is required. Instead, the functionality can be provided by appropriate tooling, which uses information from the DDI Categories to construct a dynamic classification index.

4.2.7   Classification Item

DDI Item Mapping	
DDI 3.2 Item Type	Code and Category
Item Specialization	none

GSIM Property	DDI 3.2 Mapping
Code	Code -> Value
Official name	Category -> CategoryName
Alternative names	Category -> CategoryName with context specified
Explanatory notes	Category -> Description
Level number	(computed)
Generated	Category -> UserAttributePair key = cm:IsGenerated
Currently valid	Category -> UserAttributePair key = cm:IsValid
Valid from	(computed from Statistical Classification Release Date) if in a floating Statistical Classification: Category -> UserAttributePair key = cm:ReleaseDate
Valid to	(computed from Statistical Classification Termination Date) if in a floating Statistical Classification: Category ->UserAttributePair key = cm:TerminationDate
Future events	(computed from successor Category -> BasedOnObject and VersionRationale)
Changes from previous version	(computed from VersionRationale)
Updates	(computed from VersionRationale)
Statistical Classification	(computed via referenced Statistical Classifications)
Parent item	(computed)
Sub items	(computed)
Linked items	(computed)
Case laws	Category -> UserAttributePair key = cm:CaseLaw type=json
Case law descriptions	Category -> UserAttributePair key = cm:CaseLaw type=json
Case law dates	Category -> UserAttributePair key = cm:CaseLaw type=json

4.2.8   Map

DDI Item Mapping	
DDI 3.2 Item Type	CategoryGroup
Item Specialization	CategoryGroup -> TypeOfCategoryGroup = cm:CategoryMapping

GSIM Property	DDI 3.2 Mapping
Source item	CategoryGroup -> DefiningCategoryReference
Target item (s)	CategoryGroup -> CategoryReference

Partial/complete	CategoryGroup -> UserAttributePair key = cm:IsComplete
Valid from	if in a floating Statistical Classification: CategoryGroup -> UserAttributePair key = cm:ReleaseDate
Valid to	if in a floating Statistical Classification: CategoryGroup ->UserAttributePair key = cm:TerminationDate

4.2.9   Classification Index Entry

DDI Item Mapping	
DDI 3.2 Item Type	Category
Item Specialization	none

GSIM Property	DDI 3.2 Mapping
Text	Category -> Label
Statistical Classification	(computed via parent Statistical Classification)
Codes	(computed via parent Statistical Classification)
Valid from	if in a floating Statistical Classification: Category -> UserAttributePair key = cm:ReleaseDate
Valid to	if in a floating Statistical Classification: Category ->UserAttributePair key = cm:TerminationDate
Coding instructions	Category -> UserAttributePair key = cm:CodingInstructions

5   CONTROLLED VOCABULARIES

5.1   CONTROLLED VOCABULARY FOR TYPES OF CODELISTGROUPS

Value of the Code	Descriptive Term of the Code
cm:ClassificationSeries	Classification Series
cm:ClassificationFamily	Classification Family

5.2   CONTROLLED VOCABULARY FOR TYPES OF CATEGORYGROUPS

Value of the Code	Descriptive Term of the Code
cm:CorrespondenceTable	Correspondence Table
cm:CategoryMapping	Category Mapping

5.3   CONTROLLED VOCABULARY FOR USERATTRIBUTEPAIR KEYS

Value of the Code	Descriptive Term of the Code
cm:Context	Classification Series Context
cm:UnitsClassified	Objects and Units Classified

cm:Owner	The authority which maintains the Classification Series or Correspondence Table
cm:Introduction	Statistical Classification introduction
cm:ReleaseDate	Release date
cm:TerminationDate	Termination date
cm:IsCurrent	Current
cm:MaintenanceUnit	Maintenance unit
cm:ContactPerson	Contact persons
cm:LegalBase	Legal base
cm:Copyright	Copyright
cm:IsDisseminationAllowed	Dissemination allowed
cm:IsVersion	Is the Statistical Classification a Version
cm:IsUpdate	Is the Statistical Classification an Update
cm:IsFloating	Is the Statistical Classification Floating
cm:Predecessor	Statistical Classification Predecessor
cm:Successor	Statistical Classification Successor
cm:DerivedFrom	Statistical Classification which was Derived from
cm:UpdatesPossible	Are updates possible to the Statistical Classification
cm:VariantPurpose	Purpose of variant
cm:SourceClassification	The DDI 3.2 URN of the source Statistical Classification in the Correspondence Table
cm:TargetClassification	The DDI 3.2 URN of the target Statistical Classification in the Correspondence Table
cm:SourceLevel	Level of source Statistical Classification being related
cm:TargetLevel	Level of target Statistical Classification being related
cm:IsGenerated	Level was automatically generated
cm:IsValid	The Classification Item is currently valid
| **cm:CaseLaw** | Law rulings related to a Classification Item. Content uses the following json format:<br><br>```json
{
    "Title": {
        "en": "Example name",
        "de": "..."
    },
    "Description": {
        "en": "Description of the case law",
        "de": "..."
    },
    "Date": "2009-04-12T20:44:55"
}
``` |
| **cm:IsComplete** | Specifies if the Map is complete or partial |
| **cm:CodingInstructions** | Additional information about the coding process |

6   REFERENCES

1. United Nations Economic Commission for Europe (UNECE), on behalf of the international statistical community. Generic Statistical Information Model (GSIM): Statistical Classifications Model Version 1.1. December 2013.
2. DDI Alliance. DDI Lifecycle 3.2 standard. http://www.ddialliance.org/Specification/DDI-Lifecycle/3.2/. March 12, 2014.
3. Neuchâtel Terminology Model: Classification database object types and their attributes. Version 2.1 http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=14319930
4. OECD Glossary of Statistical Terms

7   ADDITIONAL INFORMATION

For additional information regarding the topics discussed in this paper, please contact the authors.

8   AUTHORS

Corresponding Authors:

Jeremy Iverson	Colectica
Mogens Nielsen	Statistics Denmark
Dan Smith	Colectica

The following people contributed to this document.

Klas Blomqvist	Statistics Sweden
Flemming Dannevang	Statistics Denmark
Jeremy Iverson	Colectica
Jannik Jensen	Danish Data Archive
Kim Jorgensen	Statistics Denmark
Mogens Nielsen	Statistics Denmark
Dan Smith	Colectica
Riley White	Statistics Denmark

9   ACKNOWLEDGEMENTS

10 APPENDIX 1: EXAMPLE OF SIC 2007

Forthcoming in version 1.1