

THEMA Global Subject Scheme

Definition of scheme composition and code structures, version 0.9

Thema Technical Committee, April 2013

Purpose

This document accompanies the Pilot Draft (v0.9) of the *Thema* Global Subject Scheme, and is intended to define the overall composition of the scheme; whether elements are mandatory, repeatable; the structure of codes, code lengths, etc. This is to aid systems developers in preparing to accommodate *Thema* codes, even though the content of the scheme itself may change (within the parameters described here).

Composition

The *Thema* scheme comprises the following elements:

- Subject Categories
- Geographical Qualifiers
- Language Qualifiers
- Time Period Qualifiers
- Educational Purpose Qualifiers
- Interest Age & Special Interest Qualifiers
- Style Qualifiers

Of these elements, only a Subject Category is mandatory for the classification of a book product record; Qualifiers are used with Subject Categories as required.

Each Subject Category or Qualifier consists of a code and a heading. The code is universal; the heading may be presented in any language (the *Thema* scheme is published and maintained in English, though may be translated for use in any language market under the authority of the *Thema* National Group for that country or language group). In product records and message formats (such as ONIX), only the code is required to be communicated, which can then be translated by recipients as appropriate.

The set of Subject Categories is universally applicable and common to all language editions of the *Thema* scheme. The Qualifiers contain “Shared Values” which are similarly common to all editions, but may also contain “National Extensions” which typically relate to very specific topics identified as important in certain countries, but presumed not to be of universal interest or applicability. All National Extensions are available for use by all users in all countries, but are not expected to be used extensively outside the country or language group in question. Recipients of a National Extension Qualifier that they do not recognise must “resolve” the code into a broader Shared Value that is meaningful to them (see **Resolution** below).

There is no defined upper limit of the number of Subject Category values or Qualifier values that may be assigned to an individual book product record, but it is expected that a maximum of 10 of each type would sufficiently cover all reasonable circumstances. Systems designers working with systems which require limits to be placed on data element lengths and/or number of occurrences are strongly advised to provide for the full length of codes and recommended maximum number of occurrences defined in this document.

Subject Categories

Definition:

A set of codes and headings which define the subject of book content; the codes are hierarchically arranged to indicate the relationship between the headings, from broad to specific terms. Codes are extended hierarchically on a "one character, one level" basis.

Code Format:

Variable length alpha-numeric, 1-8 characters, in a single string, according to the formula:

A + <3A + <4n

where:

A is a single upper-case alpha character [A-Z]; minimum length

+ indicates the distinct elements of the code formula only; is not part of the code

<3A is up to 3 further upper-case alpha characters [A-Z]

<4n is up to 4 numeric digits other than zero [1-9]

Thus: **AAAAnnnn**

Occurrence:

One mandatory occurrence per record, up to nine further option occurrences may be expected.

Examples (with headings in English):

A	The Arts
FGH	Historical adventure fiction
QRFB23	Zen Buddhism

Qualifiers - General

Definitions and Examples of each type of Qualifier follow, but each Qualifier code conforms to the following structure, designed to accommodate both shared values and National Extensions:

Code Format

Variable length, alpha-numeric, 2-19 characters, according to the formula form:

n + <8A [+ / + CC + . + <6A]

where:

n is a numeric digit [currently 1-6 though use of further digits is reserved]

+ indicates the distinct elements of the code formula only; is not part of the code

<8A is between 1 and 8 upper-case alpha characters; **nA** is the minimum length for a shared Qualifier code

All shared Qualifier codes comprise these first two elements; the following elements [in brackets] are used only where National Extensions apply:

/ is the punctuation mark slash "/", indicating start of National Extension but with no meaning in itself (NB it is *not* used to mark the *end* of a Shared Value and is not used if no further characters indicating a National Extension are present)

CC is a two-letter Country Code indicating the source country (or group) of a National Extension (eg DE = Germany). See Notes below.

. is the punctuation mark full stop or period “.”, indicating the start or end of a country code. Will always be followed by further alpha characters (NB it is *not* used to mark the *end* of a country code and is not used if no further characters are present); the minimum length of a National Extension is **nA/CC**, though use of subsequent letters to indicate detail of National Extension is typical and recommended

<6A is between 1 and 6 alpha characters, indicating detail of National Extensions in hierarchical and sequential relationships, as required by the National Group submitting the value.

Thus: **nAAAAAAAA/CC.AAAAA**

Notes

It is arguably redundant to use a full stop or period to indicate the end of Country Codes as they always comprise two letters and always follow a slash. However, the two-letter Country Code represents the only exception to the “one character, one level” hierarchical structure in *Thema* coding, so it has been decided to demarcate them explicitly to break up the potentially long sequences of letters that comprise National Extensions to aid visual clarity and parsing.

Geographical Qualifiers

Definition:

A set of codes and headings which define the geographical scope or applicability of book content (such as the location of a travel guide, the setting of a novel, etc); the codes are hierarchically arranged to indicate the relationship between the headings, from broad levels (eg continents) to specific terms (eg countries or major geophysical features). Geographical Qualifiers contain Shared Values and National Extensions.

Code Format:

As for all Qualifiers above, but initial numeric prefix = 1.

Occurrence:

Optional, as required; up to ten occurrences may be expected.

Examples:

1H	Africa
1DDF	France
1MBZTD	Great Barrier Reef
1KBC/CA.ASF	Southern Alberta: Edmonton Capital Region

Language Qualifiers

Definition:

A set of codes and headings which define the language(s) to which the book content applies (such as a linguistics work, phrasebook or dictionary, etc), as distinct from the language of the text itself (ie the it indicates the language a book is about not the language it is written in). The codes are hierarchically arranged to indicate the relationship between the headings, from broad levels (eg language groups or families) to specific terms (eg languages or dialects). Language Qualifiers currently contain only Shared Values, though the right to add National Extensions is reserved.

Code Format:

As for all Qualifiers above, but initial numeric prefix = 2.

Occurrence:

Optional, as required; up to ten occurrences may be expected.

Examples:

2B	Indic, East Indo-European & Dravidian languages
2HCBD	kiSwahili (Swahili)

Time Period Qualifiers

Definition:

A set of codes and headings which define the time period range of book content (eg a history book, memoir, etc); the codes are hierarchically arranged to indicate the relationship between the headings, from broad periods to specific terms (eg decades). Time Period Qualifiers contain Shared Values and National Extensions.

Code Format:

As for all Qualifiers above, but initial numeric prefix = 3.

Occurrence:

Optional, as required; up to ten occurrences may be expected.

Examples:

3M	c 1500 onwards to present day
3MPBGJ/ES.B	España: Guerra civil (1936 - 1939)

Educational Purpose Qualifiers

Definition:

A set of codes and headings which define the curriculum, examination or level for which educational material is specifically designed (such as school textbooks, language learning coursework, etc); the codes are hierarchically arranged to indicate the relationship between the headings. Educational Purpose Qualifiers contain Shared Values and discrete sequences of National Extensions.

Code Format:

As for all Qualifiers above, but initial numeric prefix = 4.

Occurrence:

Optional, as required; up to ten occurrences may be expected.

Examples:

4GB	For International Baccalaureate (IB) Diploma
4Z/AA	For Arab world educational systems
4Z/UK.SD	For Scottish Curriculum National 5

Interest Age & Special Interest Qualifiers

Definition:

A set of codes and headings which define the particular reading age or level, groups of people or events for which book content is specifically intended (such as children's books, material for specific public holidays, etc); the codes are hierarchically arranged to indicate the relationship between the headings. Special Interest & Interest Age Qualifiers contain Shared Values and National Extensions.

Code Format:

As for all Qualifiers above, but initial numeric prefix = 5.

Occurrence:

Optional, as required; up to ten occurrences may be expected.

Examples:

5AQ	Interest age: from c 14 years
5HKU	Engagement / Wedding / Marriage
5PG/US.H	Latin / Hispanic American interest

Style Qualifiers

Definition:

A set of codes and headings which define the particular style of artistic or creative expression covered by book content (such as books on art history, music, literary studies, etc); the codes are arranged in flat lists in broadly alphabetical sequence (based on standard English spellings). Style Qualifiers currently contain only Shared Values, though the right to add National Extensions is reserved.

Code Format:

As for all Qualifiers above, but initial numeric prefix = 6.

Occurrence:

Optional, as required; up to ten occurrences may be expected.

Examples:

6AB	Baroque
-----	---------

Resolution

Given that the *Thema* Qualifiers contain National Extension codes that are not expected to be used by all participants, in order for universal applicability to be maintained, a procedure of Resolution is defined as follows.

Any recipient of a National Extension code (clearly indicated by the */CC* structure) not known to or used by that recipient should resolve the code to its closest Shared Value precedent, simply by ignoring all characters including and following the slash. This ensures that, although some detail is lost, a level of meaningful and pertinent information can be communicated to all users.

Thus a recipient not using the Swedish National Extensions, for example, can resolve them as follows:

3KH/SE.H Sverige: Vikingatiden (c 800 - c 1050)

Is resolved to:

3KH c 500 CE to c 1000 CE

And

1DNS/SE.CH Norrbotten

Is resolved to:

1DNS Sweden

This Resolution process can be implemented as an automatic procedure.

Entry and Display of Thema values on Systems

The *Thema* scheme is implemented on a range of systems and bibliographic record formats over which it has no control. The Category and Qualifier values may be held in a variety of field structures so long as two principles are adhered to:

- 1) the beginning and end of each code value (Category or Qualifier) must be clearly demarcated, ie they can be held as a string in a single field so long as there are delimiters between each value;
- 2) the Qualifiers are associated with the record as a whole, and should/need not be explicitly linked to a specific Category.

It is recommended that, at minimum, all Category values are held separately from all Qualifier values. If multiple values are held in the same field, the use of delimiters such as semi-colon “;” is required. Slash “/” and full stop or period “.” may NOT be used as delimiters as they are used within Qualifier code structures.

For clarity, although the Subject Category codes are alphabetical and hierarchical, and the Qualifiers begin with a number from 1 to 6, it is not essential to display the sections, or indeed the headings in alphabetical or sequential order in all contexts; they can be arranged to suit the user as required.

Use of Thema Categories and Qualifiers in ONIX

Thema is recognised in ONIX, and can be sent as part of any ONIX 2.* and ONIX 3.* messages, using standard ONIX practice for subject classification metadata. The relevant codes are values 93-99 in Code Lists 26 and 27.

Table of Code Structures

Element	Code begins	May contain	Length	Mandatory / Optional	Occurrence
Categories	A-Y	A-Z 1-9	1-8	Mandatory	1-10
Geographical Qualifiers	1	1 A-Z / .	2-19	Optional	0-10
Language Qualifiers	2	2 A-Z / .	2-19	Optional	0-10
Time Period Qualifiers	3	3 A-Z / .	2-19	Optional	0-10
Educ Purpose Qualifiers	4	4 A-Z / .	2-19	Optional	0-10
Interest Qualifiers	5	5 A-Z / .	2-19	Optional	0-10
Style Qualifiers	6	6 A-Z / .	2-19	Optional	0-10