



## NewsCodes notation in NewsML 1 TopicSets

Document history				[Document URN: urn:iptc:workdoc:ncd:0507:1 ]
Revision	Issue Date	Pages	Author (revised by)	Remark
1	2005-04-08	4	Michael Steidl	

### 1 Overview

- The IPTC NewsCodes are sets of metadata values.
- Each NewsCode represents a certain concept of the real world.
- The different sets of NewsCodes represent different scopes of concepts.
- A certain single NewsCode pertains to only one set of NewsCodes.
- As they are published, revised and re-published by the IPTC as standardisation body NewsCodes have to be considered as continuously evolving “controlled vocabularies”.

### 2 NewsCodes structures

#### 2.1 Set of information to define a single NewsCode

To define a single NewsCode this set of information is required:

(Note: consider the terms in all uppercase below as proper names reused throughout this document)

- **CODE** → a string of alphanumerical characters which should be an abstract reference to the concept. In practice this CODE could refer to the concept in human language but it has not to.
- **NAME** → a string of (not more than 255) alphanumerical characters = a short name for the represented concept in a certain human language
- **EXPLANATION** → a string of (not more than 255) alphanumerical characters = a comprehensive explanation of the represented concept in a certain human language.

The properties below pertain to a NewsCode’s membership to a set of NewsCodes:

- **FIRSTVERSION** → positive integer = version of a certain set of NewsCodes this NewsCode was added to.
- **CHANGEVERSION** → positive integer = version of a certain set of NewsCodes this NewsCode was changed by.
- **CHANGECOMMENT** → a string of (not more than 255) alphanumerical characters = optional information about the reason for any change of this NewsCode.
- **DEPRECATEDVERSION** → positive integer = version of a certain set of NewsCodes this NewsCode was deprecated by = this code should not be assigned to data anymore, but old references to this code don’t become invalid.

#### 2.2 Set of information about a set of NewsCodes

The different sets of NewsCodes are identified by NewsCodes themselves – so all of the above applies to a set of NewsCodes as well:

Each set of NewsCodes has a CODE, a NAME, and an EXPLANATION.

Furthermore each set of NewsCodes is under versioning control. The version is expressed by a positive integer starting with 1 and incremented by 1 for each change.



## 2.3 Intrinsic structures of certain sets of NewsCodes

In general the CODE representing a concept is an alphanumeric string without any further relevance to structures of a set of NewsCodes.

But currently for three sets of NewsCodes an internal structure is defined and this structure is expressed – or implemented – by a structure for the CODE.

### 2.3.1 Subject and SubjectQualifier NewsCodes

These sets of NewsCodes represent a three level structure:

- Top level = “Subject”
- Second level = “Subject Matter”
- Third level = “Subject Detail”
- Each NewsCode is present in a treelike structure:
  - All “Subject” NewsCodes relate to an abstract (non-existent) tree root as parent.
  - A “Subject Matter” NewsCode always relates to one “Subject” NewsCode as parent.
  - A “Subject Detail” NewsCode always relates to one “Subject Matter” NewsCode as parent.

The structure below was chosen to express these levels and relationships in the CODE:

- The CODE is exactly 8 characters long
- The CODE is made of digits (=numeric characters) only
- The numbering sequence of a CODE’s characters used below is from left to right
- Digits 1 and 2 represent the “Subject” level
- Digits 3, 4, and 5 represent the “Subject Matter” level
- Digits 6, 7, and 8 represent the “Subject Detail” level
- A “Subject” must never be represented by the digits “00”.
- For a NewsCode of the “Subject” level its CODE’s digits 3 through 8 are “0”
- A “Subject Matter” or a “Subject Detail” must never be represented by the digits “000”.
- For a NewsCode of the “Subject Matter” level its CODE’s digits 6 through 8 are “0”
- Subject Qualifier NewsCode’s CODE digits 3,4 and 5 are always set to “000”

Explanation/example:

- The CODE for the “Arts”-NewsCode of level “Subject” is “01000000”
- The CODE for the “Cinema”-NewsCode of level “Subject Matter” is “01005000” – it indicates the “Arts”-NewsCode as parent by having “01” as Subject-digits.
- The CODE for the “Film Festival”-NewsCode of level “Subject Detail” is “01005001” – it indicates the “Cinema”-NewsCode as parent by having “01005” as Subject- and Subject Matter-digits.

### 2.3.2 Scene NewsCodes

This set of NewsCodes represents a three tier and a two level structure:

- Top tier = “Media Type”. No NewsCodes of this level are defined, see below.
- Second tier = “Scene”, represents also the top level.
- Third tier = “Scene Detail”, represents the second level.
- Only NewsCodes of the “Scene” and “Scene Detail” level are defined. “Media Type” can be considered as kind of a qualifier.

The structure below was chosen to express these tiers in the CODE:

- The CODE is exactly 6 characters long
  - The CODE is made of digits (=numeric characters) only
  - The numbering sequence of a CODE’s characters used below is from left to right
  - Digits 1 and 2 represent the “Media Type” tier
- Currently only the sequence “01” is defined for photos.



- Digits 3 and 4 represent the “Scene” tier
- Digits 5 and 6 represent the “Scene Detail” tier
- “Media Type” digits must never be “00”.
- A “Scene” must never be represented by the digits “00”.
- Each defined NewsCode has non-zero digits for the “Media Type” and “Scene” tier.

Explanation/example:

- The CODE for the “Headshot”-NewsCode of level “Scene” is “010100”, the leading digits “01” indicate it is a Scene NewsCode for photos.

## **3 NewsCodes notation as a NewsML 1 TopicSet**

### **3.1 Introduction**

The information about a single NewsCode and about a set of NewsCodes is defined at a level agnostic to implementation technologies.

A representation of a NewsCode and sets of NewsCodes can be implemented in different technical ways.

IPTC currently provides the NewsCodes by:

- HTML pages as human readable lists (see: [www.newscodes.org](http://www.newscodes.org) )
- NewsML 1 TopicSet files (downloadable from this site)

### **3.2 NewsML 1 TopicSet files.**

Introductory note: NewsML is an XML application. Details on NewsML 1 TopicSets can be found in chapter 8 of the “NewsML 1.2 Guidelines” document, available from the NewsML website [www.newsml.org](http://www.newsml.org)

#### **Short overview:**

- A single NewsCode is expressed by a <Topic>
- A set of NewsCodes is expressed by a <TopicSet> comprising several <Topic>s
- A NewsCode’s CODE is expressed by the <FormalName> of a <Topic>
- A NewsCode’s NAME is expressed by a <Description> element of a <Topic> and its Variant attribute having a value of “Name”. Additionally the “certain human language” is expressed by an xml:lang attribute.
- A NewsCode’s EXPLANATION is expressed by a <Description> element of a <Topic> and its Variant attribute having a value of “Explanation”. Additionally the “certain human language” is expressed by an xml:lang attribute.
- A NewsCode’s CHANGECOMMENT is expressed by a <Description> element of a <Topic> and its Variant attribute having a value of “ChangeComment”. Additionally the “certain human language” is expressed by an xml:lang attribute.
- A NewsCode’s FIRSTVERSION is expressed by a <Description> element of a <Topic> and its Variant attribute having a value of “FirstVersion”.
- A NewsCode’s CHANGEVERSION is expressed by a <Description> element of a <Topic> and its Variant attribute having a value of “ChangeVersion”.
- A NewsCode’s DEPRECATEDVERSION is expressed by a <Description> element of a <Topic> and its Variant attribute having a value of “DeprecatedVersion”.

**Example:**

Excerpt from the Scene NewsCodes TopicSet file:

```

...
<TopicSet Duid="iptc.scene" Scheme="IptcTopicType" FormalName="Scene">
  <Topic>
    <TopicType Scheme="IptcTopicType" FormalName="Scene" />
    <FormalName Scheme="IptcScene">012100</FormalName>
    <Description xml:lang="en-GB" Variant="Name">posing</Description>
    <Description xml:lang="en-GB" Variant="Explanation">Subject or subjects posing such
as a "victory" pose or other stance that symbolizes leadership.</Description>
    <Description xml:lang="en-GB" Variant="ChangeComment">none</Description>
    <Description xml:lang="en-GB" Variant="ChangeVersion">1</Description>
    <Description xml:lang="en-GB" Variant="FirstVersion">1</Description>
    <Description xml:lang="en-GB" Variant="DeprecatedVersion">0</Description>
  </Topic>
  <Topic>
    <TopicType Scheme="IptcTopicType" FormalName="Scene" />
    <FormalName Scheme="IptcScene">012000</FormalName>
    <Description xml:lang="en-GB" Variant="Name">performing</Description>
    <Description xml:lang="en-GB" Variant="Explanation">Subject or subjects on a stage
performing to an audience</Description>
    <Description xml:lang="en-GB" Variant="ChangeComment">none</Description>
    <Description xml:lang="en-GB" Variant="ChangeVersion">1</Description>
    <Description xml:lang="en-GB" Variant="FirstVersion">1</Description>
    <Description xml:lang="en-GB" Variant="DeprecatedVersion">0</Description>
  </Topic>
.....
</TopicSet>
...
=== END of document ===

```