



NewsCodes notation in NewsML 1 TopicSets

Document history				[Document URN: urn:iptc:workdoc:ncd:0507:2]
Revision	Issue Date	Pages	Author (revised by)	Remark
1	2005-04-08	4	Michael Steidl	
2	2005-06-01	4	Michael Steidl/John Minting	wording reviewed, changes to 2.3.2. Scene NC explanation

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 the standardisation body NewsCodes have to be considered as continuously evolving “controlled vocabularies”.

2 NewsCodes structures

2.1 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 .
- **NAME** → a string of (not more than 255) alphanumerical characters = a short name for the concept in a specific human language
- **EXPLANATION** → a string of (not more than 255) alphanumerical characters = a comprehensive explanation of the concept in a specific human language.

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

- **FIRSTVERSION** → positive integer = version of the set of NewsCodes this NewsCode was first added to.
- **CHANGEVERSION** → positive integer = version of the set of NewsCodes in which this NewsCode was amended.
- **CHANGECOMMENT** → a string of (not more than 255) alphanumerical characters = optional information concerning the reason for any change to this NewsCode.
- **DEPRECATEDVERSION** → positive integer = version of the set of NewsCodes in which this NewsCode was deprecated = this code should not be assigned in the future , but previous references to this code do not become invalid.

2.2 Information about a set of NewsCodes

The sets of NewsCodes are listed and identified in another NewsCodes set – thus all of the information as specified in 2.1 also applies to a set of NewsCodes:

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 revision.



2.3 Intrinsic structures of certain sets of NewsCodes

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

Currently three sets of NewsCodes have an internal structure 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 employ 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:

- Each CODE consists of 8 numeric characters
- 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”.
- A “Subject” level NewsCode will have digits 3 through 8 as “0”
- A “Subject Matter” or a “Subject Detail” must never be represented by the digits “000”.
- A “Subject Matter” level NewsCode will have digits 6 through 8 as “0”
- Subject Qualifier NewsCode’s will always have digits 3,4 and 5 set to “0”

Explanation/example:

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

2.3.2 Scene NewsCodes

This set of NewsCodes employs a two tier structure:

- Top tier = “Media Type”. see below.
- Second tier = “Scene Codes”, split into two levels:
 - Top level = “Scene”
 - Second level = “Scene Detail”, adding additional granularity to the “Scene” level
- Only NewsCodes of the “Scene” or “Scene Detail” level are defined with a “Media Type” assigned to it, hence the “Media Type” can be considered as a qualifier.

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

- The CODE consists of 6 numeric characters
 -
 - 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 “01” is defined for photos.



- Digits 3 and 4 represent at the “Scene Code” “Scene” level
- Digits 5 and 6 represent at the “Scene Code” “Scene Detail” level
- “Media Type” digits must never be “00”.
- A “Scene” must never be represented by the digits “00”.

Explanation/example:

- The CODE for the “Headshot”-NewsCode is “010100”, the leading digits “01” indicate it is a Scene NewsCode for photos, with no “Scene Detail” as digits 5 and 6 are “00”.

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)
- 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

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 “ specific 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 “ specific 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 “ specific 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 ===

```