



# IPTC

---

*IPTC - NAA*  
*Subject*  
*Reference*  
*Scheme*

[Comité International des Télécommunications de Presse](#)

This document provides background on development of the revised Subject Reference scheme and guidelines for using it. The new coding structure is designed to be used within the IIM (via new DataSets); within the NITF (elements and attributes) and within NewsML (attribute values drawn from a TopicSet); or in any situation where news material needs to be categorised. These are the main points:

1. Theory and Structure
  - Object Types
  - Attributes
  - Subject References
  - Synonyms
  - Qualifiers
2. Reference Mechanisms
3. IPTC TopicSets
4. Changes to the Subject Reference Scheme
5. Implementations and Ad Hoc Extensions
6. ProviderID Registration
7. Legacy Format Issues
8. Lists of Subject Names
9. Subject Viewer

### 1. Theory and Structure

The IPTC Category Code Working Party devised a new hierarchy to describe the content of news material; one part of it is called "Subject." We made a deliberate break with the current terminology, to avoid confusion with "Category" and "Subcategory" in the existing transmission formats. Use of the new subject reference system must be considered in the context of the descriptive hierarchy.

First, we describe all individual news items as objects, whether they are text, graphics, photos, audio, video or some combination of these. Each object in turn has a type. This can be compared with CLASS in the object-oriented theory. We have identified three different object types:

- News (the default; most objects will be "news")
- Data (intended to be used for tabular text files, such as statistics or market lists, as opposed to story text)
- Advisory (news service messages, generally not published)
- Maintenance (A system-to-system message that is related to communication channels)
- TopicSet (A set of FormalNames and their meanings, changes to which are managed in a controlled manner, used in support of NewsML)
- DTD (An XML Document Type Definition)
- Document (A non-news document)
- Catalog (A container for NewsML Resource and TopicUse Elements)
- Alert (Preliminary warning message about an important breaking news event)

Second, we concluded that some items in the existing Category Codes are not truly topics in and of themselves, but rather are concepts that might apply to news material in any category. These we call attributes. An attribute describes the nature or characteristics of a news object, not specifically its content. The notion of attribute is valid for all Object Types. The new plan contains the following attributes:

Attribute No.	Object Attribute Name	Object Attribute Description
001	• Current	Indicates the object content is about events taking place at the time of the report.
002	• Analysis	The object contains data and conclusions drawn by a journalist who has researched the story in depth.
003	• Archive	The object contains material distributed previously that has been

## DRAFT AUG 2001

	material	selected from the originator's archives.
004	• Background	The object provides some scene setting and explanation for the event being reported.
005	• Feature	The object content is about a particular event or individual that may not be significant to the current breaking news.
006	• Forecast	Used when the object contains opinion as to the outcome of a future event.
007	• History	The object content is based on previous rather than current events.
016	• Interview	The object contains a report of a dialogue with a news source that gives it significant voice (includes Q & A).
008	• Obituary	The object contains a narrative about an individual's life and achievements for publication after his or her death.
017	• From the Scene	The object contains a report from the scene of an event.
009	• Opinion	The object contains an editorial comment that reflects the views of the author.
010	• Polls & Surveys	The object contains numeric or other information produced as a result of questionnaires or interviews.
022	• Press Release	The object contains promotional material or information provided to a news organisation.
011	• Profile	The object contains a description of the life or activity of a news subject (often a living individual).
012	• Results Listings & Statistics	The object contains alphanumeric data suitable for presentation in tabular form.
018	• Retrospective	The object contains material that looks back on a specific (generally long) period of time such as a season, quarter, year or decade.
013	• Side bar & Supporting information	<b>The object contains a related story that provides additional insight into the news event being reported.</b>
014	• Summary	The object contains a a single item synopsis of a number of news stories (generally unrelated).
015	• Transcript & Verbatim	The object contains a word for word report of a discussion or briefing without significant journalistic intervention.
	•	
020	• Update	The object contains an intraday snapshot (as for electronic services) of a single news subject.
021	• Wrapup	The object contains a recap of a running story (such as the end of the day).
022	• Synopsis	The object contains a condensed version of a single news item.
023	• Quote	The object contains a one or two sentence verbatim in direct quote.
024	• Press-Digest	The object contains an editorial comment by another medium completely or in parts without significant journalistic changes.
025	• Review	The object contains a critique of a creative activity or service (for example a book, a film or a restaurant).
026	• Curtain Raiser	Used when the object contains information about the staging and outcome of an immediately upcoming event.
027	• Actuality	The object contains the recording of the event.
028	• Question & Answer	The object contains the interviewer and subject questions and answers.
029	• Music	The object contains music alone.
030	• Response to a Question	The object contains a reply to a question.
031	• Raw Sound	The object contains unedited sounds.
032	• Scener	The object contains a description of the event circumstances.
033	• Text Only	The object contains a transcription of text.
034	• Voicer	The object contains only voice.

The default value is Current. This is intended to be used when no other attribute applies (or when time

## DRAFT AUG 2001

precludes making a selection). Taking the object and attribute concepts together, then, a typical wire story would be a News Object with an attribute value of Current -- that is, it's a piece of routine text on breaking news.

The third piece of the architecture is the Subject Reference, which identifies the general content of the News Object. The Subject itself is at the top of a three-level hierarchy: the second level is called Subject Matter and the third level Subject Detail. To date, we have developed Subject Matter lists for all 17 broad topics; and Subject Detail for Economy, Business & Finance, Politics and Sport.

The 17 top-level Subjects (and their three-letter abbreviations) are below; the full lists of all three levels are provided elsewhere.

Arts, Culture & Entertainment (*ACE*)  
Crime, Law & Justice (*CLJ*)  
Disasters & Accidents (*DIS*)  
Economy, Business & Finance (*FIN*)  
Education (*EDU*)  
Environmental issues (*ENV*)  
Health (*HTH*)  
Human Interest (*HUM*)  
Labour (*LAB*)  
Lifestyle & Leisure (*LIF*)  
Politics (*POL*)  
Religion and Belief (*REL*)  
Science & Technology (*SCI*)  
Social Issues (*SOI*)  
Sport (*SPO*)  
Unrest, Conflicts & War (*WAR*)  
Weather (*WEA*)

A few important notes about using the Subjects:

- They are optional. No one in the news business wants to impede rapid delivery of breaking material while a reporter or editor ponders assignment of a proper subject reference (or indeed any other administrative task). If there is no time to apply a reference, or uncertainty which would fit best, items may move with no reference at all – as they sometimes do today. This amounts to an unspecified category of "None" or "General."
- Especially to augment identification of important news items that lack a Subject, news services will continue to indicate news importance -- "bulletin," "urgent" and the like -- to help customers find hot stories whatever the topic. The NITF contains a new, separate element for marking news importance (currently known as "urgency"). Once the initial dust has settled on a non-categorised breaking story, subject references can be chosen and applied for the duration of the cycle.
- Subject references are repeatable. Many stories span multiple areas – sports figures involved in criminal activity, as an example. News services may select more than one Subject Reference so that on the customers' end, more than one interested news department in a given operation can spot the item. The Subject Detail and Subject Matter may be used together even if one is within the hierarchy of the other.

For example:

**"Government announces new educational policy for under 5 year olds"**

This would be coded as: 11006000 (Government, policy) and 05004000 (Education, Preschooling)

Or

**"UK Government announces new scheme for dairy farmers"**

11006000 (Government, policy) 04001004 (Agriculture, Livestock Farming)

As far as possible, we have attempted to put any given topic in the Subject, Subject Matter and Subject Detail lists only once. Entries for Subject Matters (general) are not normally extended into the Subject Detail level. Reasonable minds may differ (and they have) on whether the Subject Matter "Epidemic &

## DRAFT AUG 2001

Plague," for example, belongs under the top-level Subject of "Disasters & Accidents" rather than "Health." There are other debatable placements. "Teachers Unions" happens to be under Education, not Labour. But for our purposes, it is more important that a newsworthy topic exist *somewhere* on the list, in a more or less logical spot, with its own unique reference number. If a provider or receiver keys on the reference number or the subject matter name to mark or find related items, its particular location on the hierarchical list is irrelevant.

Despite our effort to place topics in only one spot, there are some unavoidable duplicates. The most persistent example of this is under Economy, Business & Finance. Many discrete activities that generate news in their own right are also businesses: Media and Tourism are two examples. The intent is that news on Television as a medium – entertainment, news and the like – belongs under the "Arts, Culture and Entertainment" Subject Code. News about the business of television – network ownership, corporate matters – falls under Economy/Business/Finance. A story about remote Caribbean resorts fits the Tourism entry under Lifestyle & Leisure; a profile about a cruise ship company is Business. We define Sports as being "competitive exercise involving physical effort." This precludes activities of predominately mental exertion such as chess, bridge and other board or card games.

Finally, we point out the continued option of using the keyword in news transmissions. This repeatable field may be used for terms that help further define item, but that may not warrant inclusion in the subject lists. A couple examples:

- For a chess championship: Object=News, Object Type=Current, Subject=Lifestyle & Leisure, Keyword=Chess Masters.
- For a general weather story that cannot necessarily be covered by Object Type of "current": Object=News, Object Type=Feature, Subject=Weather, Keyword=Heat wave.
- Synonyms Specific institution or national organisation names are translated into the terms used in the respective language but reflect back to a single English term and the related code. A particular example of this is Parliament that has different terminology depending on the language and culture. Regional chambers should be subject to local IPR but in turn refer back to Parliament at the Subject Matter level. Some national legislative bodies (groups elected for law making or legislative power) occur at different levels within societies. Local additions should be made as subject details with local IPRs.  
Synonyms only apply at the Subject Matter level. That is, the same code is used for terms meaning the same or equivalent thing in different languages. There may be regional authorities in some countries such as Germany and France which might also be included within the generic subject matter description.

**Qualifiers** It may be necessary to provide additional information about the subject that is not covered by the first selected code. These extra terms are known as Qualifiers and serve to provide a narrower context for the news item. Suitable terms have been generated (particularly for Sports Organisations and Events). Qualifying details may not be used alone and must be related to an additional subject reference coding.

For example in Athletics for semi-finals of 100 m Men of Olympics in Sydney

15000000 subject name = Sport

15000001 subject qualifier name = Men

15000027 subject qualifier name = Round

15005000 subject matter name = Athletics

15005001 subject detail name = 100M

Additional identification can be achieved by using a Keyword = semi-finals

In order to avoid too much redundancy the above may be encoded 15000001,15000027,15005001 plus the Keyword "semi-finals".

## 2. Reference Mechanisms

It is expected that the subject lists will be translated and used in many languages, often within the same

## DRAFT AUG 2001

news organisation (providers and customers alike). Thus there is a need for fixed anchors that can be used to identify unambiguously the same topic across all data streams. We provide two.

The first is a three-letter English code for the 17 broad categories. These generally represent the first letters of the word or words in the English version of the subject. This is designed to be used with legacy formats where long numeric strings are inappropriate.

In addition, each Subject may be referenced by a 8 digit number (01000000 to 17000000), which in turn is the basis of unique eight-digit numbers for each of the second- and third-level descriptors. The numbering scheme also allows for logical extensions to the lists.

### 3. IPTC TopicSets

The list of object types is maintained as an ITPC TopicSet termed `NewsItemTypes`. These `NewsItemTypes` are recommended to all NewsML originators in the news domain to ensure interoperability.

The list of object attributes is maintained as an ITPC TopicSet termed `Genre`. This TopicSet is based on the experience of news accumulated by ITPC Members. However, NewsML users may choose to produce their own TopicSet for genre based on their business needs. The term genre is more related to a description of the news object content rather than how it might be rendered, **Accordingly the term "Sidebar & Supporting Information" is not included in the Genre TopicSet as it more properly falls within the scope of a Role within NewsML.** The entry is highlighted in red for this reason although outside of NewsML it is a valid object attribute.

The ITPC `SubjectCode` TopicSets have been developed after considerable analysis and discussion within the news industry. The structure of NewsML contains explicit elements to contain `Subject`, `SubjectMatter`, `SubjectDetail` and `SubjectQualifier`. Users who elect not to take advantage of the work behind these data can only use these named elements for their own scheme providing it conforms to the structure of the ITPC Subject Reference Scheme. If a provider develops his own Subject scheme the `Scheme` attribute must not be set equal to `IptcSubjectCodes`. (See section 5.)

The official published versions of the Subject vocabulary will be in English. There is no relevance in the sequence of numbers allocated for each part of the Subject references. They were derived initially on an alphabetical sort of the original subject matters and do not contain an implied hierarchy. However, the first two digits are always the same for a given Subject and the digits at positions 3, 4 and 5 are the same for a given SubjectMatter. The ITPC will maintain a registry of Subject Reference Numbers, Subject, Subject Matter and Subject Detail as appropriate in other languages as submitted by members or users. Translations of the English list into other languages are available (as XML and PDF files) on the ITPC Web site, or in other forms from the ITPC.

The Qualifiers are also maintained as an ITPC TopicSet called `SubjectQualifier`.

The data will be maintained in the form of an XML file that conforms to the definition of the TopicSet element in the NewsML document type definition. A TopicSet is defined as a container for Topics. The value of the `FormalName` attribute is a formal name for the TopicSet. Its meaning and permitted values are determined by a controlled vocabulary. In practice this means that the file takes the form as shown in the following fragment.

```
<!DOCTYPE TopicSet PUBLIC " urn:newsml:iptc.org:20001006:NewsMLv1.0.dtd:1"
"../DTD/NewsMLv1.0.dtd">
<TopicSet Duid="IptcSubjectCodes.subject" FormalName="SubjectCode">
  <Comment xml:lang="en">Topics whose TopicType is Subject provide a description of the
editorial content of a NewsComponent at a high level.</Comment>
  <Comment xml:lang="en">Topics whose TopicType is SubjectMatter provide a description at a
more precise level.</Comment>
  <Comment xml:lang="en">Topics whose TopicType is SubjectDetail provide a description at a
specific level.</Comment>
  <Topic Duid="sr01000000">
```

## DRAFT AUG 2001

```
<TopicType Scheme="IptcTopicType" FormalName="Subject"/>
<FormalName Scheme="IptcSubjectCodes">01000000</FormalName>
<Description xml:lang="en" Variant="Name">Arts, Culture & Entertainment</Description>
<Description xml:lang="en" Variant="Explanation">Matters pertaining to the advancement and
refinement of the human mind, of interests, skills, tastes and emotions</Description>
</Topic>
<Topic Duid="sr01001000">
  <TopicType Scheme="IptcTopicType" FormalName="SubjectMatter"/>
  <FormalName Scheme="IptcSubjectCodes">01001000</FormalName>
  <Description xml:lang="en" Variant="Name">Archaeology</Description>
</Topic>
</TopicSet >
```

This format may be transformed into other XML, HTML or text constructs using appropriate stylesheets. These formats may be imported into (for example) MS Excel. (See also 9. Subject Viewer.)

In general, extensions to the published lists may only be used with the Scheme attribute value of IptcSubjectCodes if the change has been formally considered and agreed within the IPTC. The lists will be subject to regular review and updating as deemed necessary by IPTC members. The version for a given list will be indicated by using the appropriate RevisionID on the TopicSet NewsML PublicIdentifier urn. (This is indicated by the last numeric character in the urn string.) Users experiencing a serious recurring problem with coding news items should refer back to the IPTC for advice or to propose any amendments.

In order to be responsive to user needs a "fast track" process has been developed for extensions to the Subject Detail Lists. Subject Matter and other data will be considered through the normal IPTC Change Request process at the next scheduled IPTC meeting. Proposals from organisations outside IPTC must be sent to the Managing Director who will assist in identifying at least one current IPTC member to sponsor the request.

#### 4. Changes to the Subject Reference Scheme

##### Criterea for Inclusion

IPTC has established the following criteria for evaluating the Additions to the Subject Reference Scheme

- (1) There must be strong international interest in the topic.
- (2) A topic must generate substantial news volume.
- (3) a. Either general interest newspapers typically devote a regular section or page to this topic, or  
b. It is a typical segment of a radio or TV news broadcast, or an on-line news service.
- (4) Does current news service practice support designation of this category?
- (5) The topic is unique -- not a logical subset of some other topic.
- (6) The topic must have longevity and not be of transient interest.

##### Rules For Processing Changes

1. The Change Request should be sent to the Managing Director preferably by e-mail or if not by fax. (Form is available on the IPTC Web site.)
2. Accelerate the process (similar to ISO fast track) to 28 days for Subject Details Only.
3. A jury of 3 will consider the Change Request for Subject Details. The Chairmen of the Standards Committee will appoint the Jury with concurrence of the committee itself. (These members should have a good knowledge of the Subject Reference system its background.)
4. Change Request is circulated to all membership. Members can forward constructive objections to the Jury within the first 21 days.
5. The Managing Director Assigns numbers and progresses the Change Request and the Jury Work and

## DRAFT AUG 2001

is responsible for the publication of the results.

6. For Subject Matters an Ad hoc group should address Change Request with the submitter

### 5. Implementation and Ad Hoc Extensions

The easiest way to extend the lists as published in an ad-hoc fashion is to use an appropriate Keyword. If this proves insufficient, information providers may develop extensions of the second and third level of the subject hierarchy (Subject Matter and Subject Detail). The 17 parent Subjects as published by the IPTC and their reference number may NOT be extended or altered by individual information providers.

#### NewsML Users

A provider extending the lists must be in possession of a www domain name that is used to form the ProviderID within the NewsML urn. The provider then must identify any local extensions to the Subject Reference Scheme using a value for the Scheme attribute generated by the provider. Entries so identified allow for addition of local or regional interest items that might not find general usage by other news providers. For example, a North American agency may wish to adopt Subject Details for a particular sport that would specify various leagues or variations of the sport. A provider specialising in business information might wish to develop more subject detail on specific markets, such as commodities. Both such efforts would retain the 17 top-level Subject Codes, and could retain the secondary level as well. The eight-digit reference number accommodates any such extensions.

#### Example

MySchemeSubject TopicSet controlled vocabulary is published by the Provider in a NewsItem under their own NewsML urn. Eg urn:newsml:myscheme.org:20001212:MySchemeSubjects:1

The new entry would be published thus:

An extract from the current IptcSubjectCode file is first shown for clarity

<TopicSet>

.....

<Topic Duid="sr04015000">

<TopicType FormalName="SubjectMatter"/>

<FormalName Scheme="IptcSubjectCodes">04015000</FormalName>

<Description xml:lang="en" Variant="Name">Transport</Description>

</Topic>

<Topic Duid="sr04015001">

<TopicType FormalName="SubjectDetail"/>

<FormalName Scheme="IptcSubjectCodes">04015001</FormalName>

<Description xml:lang="en" Variant="Name">Air Transport</Description>

</Topic>

<Topic Duid="sr04015002">

<TopicType FormalName="SubjectDetail"/>

<FormalName Scheme="IptcSubjectCodes">04015002</FormalName>

<Description xml:lang="en" Variant="Name">Railway</Description>

</Topic>

<Topic Duid="sr04015003">

<TopicType FormalName="SubjectDetail"/>

<FormalName Scheme="IptcSubjectCodes">04015003</FormalName>

<Description xml:lang="en" Variant="Name">Road Transport</Description>

<Description xml:lang="en" Variant="Explanation"/>

</Topic>

<Topic Duid="sr04015004">

<TopicType FormalName="SubjectDetail"/>

<FormalName Scheme="IptcSubjectCodes">04015004</FormalName>

<Description xml:lang="en" Variant="Name">Waterway & Maritime Transport</Description>

</Topic>

## DRAFT AUG 2001

```
<!DOCTYPE TopicSet PUBLIC " urn:newsml:iptc.org:20001006:NewsMLv1.0.dtd:1"
"../DTD/NewsMLv1.0.dtd">
<TopicSet Duid="MySubjectCodes.subject" FormalName="SubjectCode">
  <Topic Duid="sr04015005">
    <TopicType FormalName="SubjectDetail"/>
    <FormalName Scheme="MySchemeSubjects">04015005</FormalName>
    <Description xml:lang="en" Variant="Name">Space Transport</Description>
    <Description xml:lang="en" Variant="Explanation">Travel beyond the confines of the earth's
      atmosphere</Description>
  </Topic>
  .....
</TopicSet>
```

Existing IPTC published entries are shown in black the entry in the Providers extension TopicSet is in red. Note that numbers sequence allocated for the Subject Reference and expressed as the FormalName element value of the Topic must not be a repeat of any that have been previously published by IPTC. In the example the new digits in the sequence are 005. If a new SubjectMatter is required then the last 6 digits will be those allocated by the new Scheme provider.

In the NewsML instance the following structure is used to indicate that a new Scheme is being invoked.

```
.....
<DescriptiveMetadata>
  <SubjectCode>
    <SubjectDetail Scheme="IptcSubjectCodes" FormalName="15073001"/>
    <SubjectDetail Scheme="MySchemeSubjects" FormalName="04015005"/>
  </SubjectCode>
</DescriptiveMetadata>
.....
```

### News Industry Text Format (NITF) Users

In the NITF (also an XML document but used to mark up text content of News) there are explicit elements and attributes used to identify the relevant extracts from the Subject Reference Scheme. The appropriate DTD extracts follows:

```
<!ELEMENT tobject (tobject.property*, tobject.subject*)>
<!ATTLIST tobject
  %global-attributes;
  tobject.type          CDATA          "news"
>
<!--
  Element          tobject
  Partial description      Subject code.
  Full description      Used to identify the type of the news material,
  based on a Subject Code system.

  Attribute name          id
  Description              Unique identifier for the element.
  Attribute name          tobject.type
  Description              Values are drawn from the data in the iptc-
  newsitemtype vocabulary.
-->

<!ELEMENT tobject.property EMPTY>
  <!ATTLIST tobject.property
    %global-attributes;
    tobject.property.type  CDATA          "current"
  >
```

## DRAFT AUG 2001

```
<!--
    Element                tobject.property
    Partial description     Subject code property.
    Full description       Assigns a property to the object type.

    Attribute name         id
    Description            Unique identifier for the element.
    Attribute name         tobject.property.type
    Description            Values are drawn from the data in the iptc-
genre vocabulary.
-->
```

```
<!ELEMENT tobject.subject EMPTY>
  <!ATTLIST tobject.subject
    %global-attributes;
    tobject.subject.ipr          CDATA          " IPTC "
    tobject.subject.refnum      NMTOKEN        #REQUIRED
    tobject.subject.code        CDATA          #IMPLIED
    tobject.subject.type        CDATA          #IMPLIED
    tobject.subject.matter      CDATA          #IMPLIED
    tobject.subject.detail      CDATA          #IMPLIED
  >
  <!--
    Element tobject.subject
    Partial description:Assigns subject information to news
material based on a Subject Code system.
    Full description: A three-tiered hierarchy consisting of broad
subjects, secondary topics (Subject Matter), and tertiary topics (Subject
Detail). Values are drawn from the data in the iptc-subjectcode vocabulary.

    Attribute name         id
    Description            Unique identifier for the element.
    Attribute name         tobject.subject.ipr
    Description            Information provider reference.
    Attribute name         tobject.subject.refnum
    Description            Eight-digit reference number for all three
levels.
    Attribute name         tobject.subject.code
    Description            Three-letter code for the subject type.
    Attribute name         tobject.subject.type
    Description            First-tier subject description.
    Attribute name         tobject.subject.matter
    Description            Second-tier subject description.
    Attribute name         tobject.subject.detail
    Description            Third-tier subject description.
  -->
```

In this case the relevant source of the information is provided by the IPR value of IPTC. Users creating their own extensions must use an IPR allocated to them by the IPTC. As in the NewsML case extensions may only be at the SubjectDetail Level and Reference codes must not repeat the values already assigned by the IPTC.

### Information Interchange Model (IIM) Users

The appropriate extract from the IIM follows and explains how the Subject Reference system is to be implemented within the IIM Record 2.

## DRAFT AUG 2001

### 2:12 **Subject Reference**

Repeatable. Minimum of 13 and maximum of 236 octets consisting of graphic characters. Colon ':' is only allowed as specified, the asterisk '\*' and question mark '?' are not allowed, nor are the octet values 42 and 63.

The character encoding used for this dataset must encode the colon ':' using octet value 58, and must not use this octet value for any other purpose.

The Subject Reference is a structured definition of the subject matter. It must contain an IPR (default value is "IPTC"), an 8 digit Subject Reference Number and an optional Subject Name, Subject Matter Name and Subject Detail Name. Each part of the Subject reference is separated by a colon (:). The Subject Reference Number contains three parts, a 2 digit Subject Number, a 3 digit Subject Matter Number and a 3 digit Subject Detail Number thus providing unique identification of the object's subject.

If the Subject Matter or Subject Detail is not defined then a value of 000 is used for the Subject Matter Number and/or Subject Detail Number as appropriate.

The DataSet may be repeated when the objectdata content is relevant to several subjects of news interest. It can be independent of provider and for any media form. The provider must either use the IPTC scheme or one that has been defined and published by the provider.

## DRAFT AUG 2001

The construction of the Subject Reference is as follows:

- **Information Provider Reference (IPR)**

A name, registered with the IPTC/NAA, identifying the provider that provides an indicator of the SDR content. The default value for the IPR is "IPTC" and is mandatory if the Subject Reference exists in the IPTC coding scheme.

Individual registered Information Providers may at their discretion extend the Subject Reference lists. However, they may only add to the subject matter and/or subject detail included in the IPTC lists, and must identify this by using their registered IPR. The IPTC Subject list may not be extended.

- **Subject Reference Number**

Provides a numeric code to indicate the Subject Name plus optional Subject Matter and Subject Detail Names in the language of the service. Subject Reference Numbers consist of 8 octets in the range 01000000 to 17999999 and represent a language independent international reference to a Subject. A Subject is identified by its Reference Number and corresponding Names taken from a standard lists given in Appendix H,I &J. These lists are the English language reference versions.

- **Subject Name**

The third part, if used, is a text representation of the Subject Number (maximum 64 octets) consisting of graphic characters plus spaces either in English, as defined in Appendix H, or in the language of the service as indicated in DataSet 2:135

The Subject identifies the general content of the objectdata as determined by the provider.

- **Subject Matter Name**

The fourth part, if used, is a text representation of the Subject Matter Number (maximum 64 octets) consisting of graphic characters plus spaces either in English, as defined in Appendix I, or in the language of the service as indicated in DataSet 2:135

A Subject Matter further refines the Subject of a News Object.

- **Subject Detail Name**

The fifth part, if used, is a text representation of the Subject Detail Number (maximum 64 octets) consisting of graphic characters plus spaces either in English, as defined in Appendix J, or in the language of the service as indicated in DataSet 2:135

A Subject Detail further refines the Subject Matter of a News Object. A registry of Subject Reference Numbers, Subject Matter Names and Subject Detail Names, descriptions (if available) and their corresponding parent Subjects will be held by the IPTC in different languages, with translations as supplied by members.

## DRAFT AUG 2001

<i>Minimum of 13 and maximum of 236 Octets</i>								
IPR	ES	Subject Reference Number	ES	Subject Name	ES	Subject Matter Name	ES	Subject Detail Name
1 - 32 Octets registered with IPTC for UNO	:	Eight octets assigned by the IPTC as contained in Appendices H,I & J	:	Maximum 64 octets to reference the Subject	:	Maximum 64 octets to reference the Subject Matter	:	Maximum 64 octets to reference the Subject Detail
<i>Minimum 1, maximum of 32</i>	<i>1</i>	<i>8</i>	<i>1</i>	<i>0-64 Octets</i>	<i>1</i>	<i>0-64 Octets</i>	<i>1</i>	<i>0-64 Octets</i>

### 6. ProviderID Registration

An Information Providers Reference or ProviderID may be registered with IPTC under the following circumstances:

- a. The Information Provider is a member in good standing of IPTC or
- b. The Information Provider is a registered member of the NewsML email list or a registered holder of the Information Interchange Model
- c. The requested IPR mnemonic has not been registered by any other Information Provider.

Application should be made to the Managing Director IPTC, by EMAIL or FAX indicating the Information Providers full name, the desired IPR and a text description of the Company, Publication or Service being submitted. Both the IPR string and the description must be unique for each entry. When accepted the IPR details will be published in IPTC Provider TopicSet.

### 7. Legacy Formats

For legacy formats, where no numeric coding is possible, a single term can be used since there are implicit relations between subject detail and subject matter. If the textual form only is used for subject detail, without the subject matter, then the client systems must have the capability to resolve the implied relationship. It may be necessary to use the subject matter with the subject detail to avoid ambiguity for example radio, television etc.

The following mechanisms are proposed for dealing with legacy formats.

- a. ANPA 1312

It appears unlikely that any part of the new Subject Reference scheme can be incorporated into the ANPA 1312 transmission format. This is beyond the scope of the IPTC because of the rigid format used within ANPA 1312. The matter therefore is referred to those having the greatest interest in this item, namely the NAA Wire Services Guidelines Committee. (It is noted that any Subject Reference information would probably need to be in the text of a message).

- b. IPTC 7901

This recommendation is designed to assist users who make use of the keyword line in the IPTC TEC 7901 format. It is not intended to modify the original specification or to imply that any particular implementation is mandatory.

IPTC TEC 7901 is not a tagged format and does not provide for coding of elements such as country codes or subject reference. In short, the subject code scheme can not be added as new item to the format but existing fields have to be used for implementation.

Where an IPR decides to use this system, it may be implemented within the format as follows:

## DRAFT AUG 2001

The Country Name and the Subject, Subject Matter or Detail may be placed in the keyword line. In order not to confuse the idea of subjects with the practise of keywords, which may still be used, The following alternative definitions are listed:

Subject Matter = Primary keyword  
Subject Detail = Secondary keyword  
Keyword = Unlisted Keyword

Subject = Primary keyword  
Subject Matter = Secondary keyword  
Keyword = Unlisted Keyword

The choice of allocation of value to each keyword is left to the IPR and depends on the service requirements. The sequence for placing the above format elements on the 7901 Keyword Line is as follows:

Country Name/Primary Keyword/Secondary Keyword/Unlisted Keyword

When this recommendation is followed the Primary Keyword is mandatory, Secondary and Unlisted Keywords are optional.

### 8. IPTC SUBJECT REFERENCES (Version IPTC/1)

#### a. Subjects (Top Level)

Arts, Culture & Entertainment (*ACE*) 01000000

Matters pertaining to the advancement and refinement of the human mind, of interests, skills, tastes and emotions

Crime, Law & Justice (*CLJ*) 02000000

Establishment and/or statement of the rules of behaviour in society, the enforcement of these rules, breaches of the rules and the punishment of offenders. Organisations and bodies involved in these activities.

Disasters & Accidents (*DIS*) 03000000

Man made and natural events resulting in loss of life or injury to living creatures and/or damage to inanimate objects and property.

Economy, Business & Finance (*FIN*) 04000000

All matters concerning the planning, production and exchange of wealth.

Education (*EDU*) 05000000

all aspects of furthering knowledge of human individuals from birth to death.

Environmental issues (*ENV*) 06000000

All aspects of protection, damage, and condition of the ecosystem of the planet earth and its surroundings.

Health (*HTH*) 07000000

All aspects pertaining to the physical and mental welfare of human beings.

Human Interest (*HUM*) 08000000

Lighter items about individuals, groups, animals or objects.

Labour (*LAB*) 09000000

Social aspects, organisations, rules and conditions affecting the employment of human effort for the generation of wealth or the provision of services and the economic support of the unemployed.

Lifestyle & Leisure (*LIF*) 10000000

Activities undertaken for pleasure, relaxation or recreation outside paid employment, including eating

## DRAFT AUG 2001

and travel.

**Politics (*POL*) 11000000**

Local, regional, national and international exercise of power, or struggle for power, and the relationships between governing bodies and states.

**Religion and Belief (*REL*) 12000000**

All aspects of human existence involving theology, philosophy, ethics and spirituality.

**Science & Technology (*SCI*) 13000000**

All aspects pertaining to human understanding of nature and the physical world and the development and application of this knowledge.

**Social Issues (*SOI*) 14000000**

Aspects of the behaviour of humans affecting the quality of life.

**Sport (*SPO*) 15000000**

Competitive effort involving physical effort. Organisations and bodies involved in these activities.

**Unrest, Conflicts & War (*WAR*) 16000000**

Acts of socially or politically motivated protest and/or violence.

**Weather (*WEA*) 17000000**

The study, reporting and prediction of meteorological phenomena.

### 9. Subject Viewer

To assist users a Windows based Subject Viewer has been developed. This allows detailed examination of the XML TopicSet files that have been developed to contain the Subject Reference Scheme information. Not only can the Subject References and Qualifiers be seen in detail but the associated TopicSets for Genre, NewsItemType and MediaType are also displayable. The package allows export of text files suitable for import into other applications such as spreadsheets and for the transformation of the TopicSet XML format into other formats as needed. A stylesheet is included to provide an HTML version of the files that is displayable in a web browser. The Viewer and the associated data files may be downloaded from the [iptc.org](http://iptc.org) web site.