IPTC - NAA

Subject

Reference

System

Guidelines

International Press Telecommunications Council
Comité International des Télécommunications de Presse

Version 3 / November 2003

Introduction

This document provides background and implementation guidelines for the news industry's Subject Reference System, jointly developed by the International Press Telecommunications Council and the Newspaper Association of America. The system is designed to be used in any situation where news material needs to be categorised, including:

- within NewsML, as attribute values drawn from a TopicSet; or
- within the News Industry Text Format, as elements and attributes, or
- within the Information Interchange Model, via new DataSets.

The system is an open standard that is freely available for use either with these other IPTC standards or as a stand-alone vocabulary.

TABLE OF CONTENTS

Theory and Structure

The IPTC developed a new framework to describe the content of news material, part of which is the subject. Any implementation of the Subject Reference System should take into consideration this broader descriptive hierarchy. Its components are objects, attributes, subject reference, synonyms and qualifiers.

Media Type

The highest level descriptor for a news object is its media type. Valid values are: text, graphic, photo, audio, video and animation

Object Type

Next, each individual news object has a **type**. This can be compared with CLASS in object-oriented theory. We have identified the following object types:

- **news** (default)
- **data** (intended for tables such as statistics or lists, as opposed to narrative text)
- **advisory** (content provider messages, generally not published)
- **maintenance** (system-to-system messages related to communication channels)
- **TopicSet** (a set of FormalNames and their meanings, used in support of NewsML and governed by IPTC change-control procedures)
- **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)

Attributes

We identified a number of concepts that might apply to news material in any subject area. These we call **attributes**. An attribute describes the nature or characteristic of a news object, not specifically its content. An attribute may be applied to any of the object types. Valid attributes are listed below. The reference numbers are arbitrary and have no implied significance.

Ref. No.	Attribute name	Description
001	Current	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 material	The object contains material distributed previously that has been 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 current breaking news.
006	Forecast	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 dialogue with a news source that gives it significant voice.

Ref. No.	Attribute name	Description
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 and 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 and 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 and supporting information	The object contains a related story that provides additional insight into the news event being reported.
014	Summary	The object is a collection of synopses on news items (generally unrelated). See **Synopsis.**
015	Transcript and 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 item.
021	Wrapup	The object contains a recap of a running story (such as at the end of the day).
019	Synopsis	The object contains a condensed version of a single news item.
023	Quote	The object contains a short verbatim excerpt from an individual or other source.
024	Press-digest	The object contains material by another content provider 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 probable outcome of an immediately coming event.
027	Actuality	The object contains the audio **or video** recording of the event
028	Question and answer	The object contains questions and answers between the interviewer(s) and subject(s).
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.
035	Fixture	A document with a constant reference whose content is updated periodically.

The default attribute value is Current. This is to be used when no other attribute applies, or when time precludes making a selection. Taking the object and attribute concepts together, then, a typical article would be a **News Object** with an attribute value of **Current** – that is, it's a piece of routine text on breaking news.

Subject Reference

The next piece of the architecture is the **Subject Reference**, which identifies the general content of a News Object. This is taken from a three-level hierarchy where the top level is **Subject;** the second level is **Subject Matter** and the third level is **Subject Detail**. There are 17 top-level **Subjects,** and we have developed secondary **Subject Matter** lists for each of these. To date, there are third-level **Subject Detail** lists for three Subjects: Economy, Business and Finance, Politics, and Sport. All references are controlled by a fixed eight-digit reference number described later in this paper.

The 17 top-level Subjects are below, along with and their three-letter English abbreviations, reference numbers and definitions.

Arts, Culture and Entertainment (ACE) 01000000
Matters pertaining to the advancement and refinement of the human mind, of interests, skills, tastes and emotions

Crime, Law and 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 and 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 and 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 and Leisure (LIF) 10000000
Activities undertaken for pleasure, relaxation or recreation outside paid employment, including eating 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 and 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 and War (WAR)   16000000
Acts of socially or politically motivated protest and/or violence.

Weather (WEA)   17000000
The study, reporting and prediction of meteorological phenomena.

A few important notes about using Subjects:

- All terms used will be those that are internationally recognised rather than of particular national usage. Entries will appear in lower case unless they are proper names or registered with a leading upper case or are acronyms or abbreviations. For languages other than English the terms are capitalised in accordance with common usage and dictionary practice for that language. Where singular terms are used they cover plural representation of entries and vice versa.

- They are **optional.** IPTC members are well aware that in the context of real-time information delivery, deadline pressures may prevent reporters and editors from assigning a subject reference to a news item. If there is no time to apply a reference, or uncertainty about what would fit best, items may move with no reference at all – as they sometimes do today.

- Content providers will continue to indicate **news importance** – "bulletin," "urgent" and the like – to help customers find top stories regardless of the subject, and especially in cases where a subject isn't specified. The NITF contains a separate element for marking news importance. Once the initial dust has settled on a non-categorised breaking story, subject references can be chosen and applied to subsequent news items.

- Subject references are **repeatable.** Many stories span multiple areas – sports figures involved in criminal activity, as an example. Subject Matter and Subject Detail entries may be used together, even if one is within the hierarchy of the other.

  Examples:
      "Government announces new educational policy for under 5 year olds"
      11006000 (Government, policy)
      05004000 (Education, preschooling)

      "UK Government announces new scheme for dairy farmers"
      11006000 (Government, policy)
      04001004 (Agriculture, Livestock Farming

- Providers have the option of using a **keyword** to augment subject assignments. This field (also repeatable in NITF) may be used for terms that help further define an 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 and leisure
          keyword=chess masters

  for a general weather story:
          object=news
          object type=feature
          subject=weather

keyword=heat wave

Duplication and Inheritance

Although the original intentions of IPTC in regard to the same term appearing more than once within the SRS was to limit or even prevent this, usage has shown that a term can have more than one meaning depending upon its context and therefore should be able to appear as a Subject Matter or Subject Detail multiple times but each appearance would have a totally different "explanation".

There are numerous examples of this need:
01016000 Television as Arts Culture and Entertainment, relates to the AC&E as shown on television while:
04010010 Television as Economy, Business and Finance relates to the business aspects of television, ownership, profits/losses, ad revenue or corporate matters.

Therefore as long as the context is really different, similar terms may appear multiple times under different Subjects or Subject Matters, including Subject Details as the children of different Subject Matters under the same Subject
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, which are part of Lifestyle and Leisure.

Synonyms

Synonyms are used to describe common concepts that use different terminology depending on the cultural and legal framework. 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 that might also be included within the generic subject matter description. Every attempt has been made to use cultural-neutral English where possible.

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.

Individual users are responsible for maintaining their own lists of local synonyms and for ensuring their customers have the appropriate information to interpret any local extensions to the scheme.

Qualifiers

We discovered cases where Subject References seemed to demand auxiliary terms to make the reference more precise and useful. The most compelling example is in Sports. Many sporting events are broken down into subgroups based on such things as the gender or age of participants, the number of participants (singles, two-person, four-person etc.), or the level of competition (amateur, professional, regional, national).

We call these auxiliary terms **qualifiers.** They may not be used alone, but rather must be used in conjunction with a specific Subject (top-level) reference. Each qualifier has a unique number that is used as part of the eight-digit numeric anchor for a Subject.

Reference Mechanisms

We developed the original subject lists in English but they have been translated into many languages, as demanded by news providers and their clients. So there was a need for unambiguous anchors that could be used to identify the same topic across all content, regardless of media type or provider. We provide two.

The first is a three-letter English code for the 17 top-level categories. These generally represent the first letters of the word or words in the English version of the subject. They can be used with legacy news industry formats where long numeric strings may be inappropriate.

The more important mechanism is a unique eight-digit number that is assigned to each entry in the three-level Subject hierarchy. This number is broken down as follows:

The **first two digits** indicate the top-level **Subject**. The valid values are 01 through 17. (Leading zero is mandatory.)

The **next three digits** indicate the **Subject Matter**. Default is 000, used when no Subject Matter is specified. The rest of the values (001-999) must be used in conjunction a two-digit Subject number.

The **last three digits**, when preceded by valid Subject and Subject Matter numbers, indicate **Subject Detail**. The last three digits separately are used to indicate a **qualifier** – but *only* in concert with a Subject number. The middle three digits must be 000 when providing a qualifier.

Examples:

	Subject	Subject Matter	Subject Detail
13000000	13	000	000
	Science and Technology	(none)	(none)

	Subject	Subject Matter	Subject Detail
07002000	07	002	000
	Health	Epidemic and Plague	(none)

	Subject	Subject Matter	Subject Detail
04015002	04	015	002
	Economy, Business and Finance	Transport	Railway

	Subject	Subject Matter	Subject Qualifier
15000001	15	000	001
	Sport	(none)	Men

	Subject	Subject Matter	Subject Qualifier
15000010	15	000	010
	Sport	(none)	Professional

The numbering scheme allows for logical extensions to all of these lists. There is no relevance in the sequence of numbers allocated to entries in the scheme. They were derived initially from an alphabetical sort of the original lists 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.

Controlled Vocabulary Maintenance

The various components of the Subject Reference System – such as object types, attributes and qualifiers – will be maintained through the TopicSet mechanism of the IPTC's NewsML standard.

The list of **media types** is maintained as an IPTC TopicSet termed mediatypes. These mediatypes are recommended to all content originators using NewsML in the news domain to ensure interoperability.

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

The list of **object attributes** is maintained as an IPTC TopicSet termed Genre. The values in this TopicSet are derived from the experience of news gathering organisations within IPTC. But NewsML users in other domains may choose to produce their own TopicSet for genre based on their business needs. The term "genre" is related more to a news object's content than how it might be rendered. Accordingly the term "Sidebar and 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.

Subject, Subject Matter and **Subject Detail** are managed in the NewsML TopicSet called SubjectCode. **Subject Qualifiers** are maintained as an IPTC TopicSet called SubjectQualifier**.**

The IPTC 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 IPTC Subject Reference System. If a provider develops his own Subject scheme the Scheme attribute must not be set equal to IptcSubjectCodes. (See section 5.)

The IPTC's official published versions of the Subject vocabularies will be in English. The IPTC will maintain a registry of corresponding lists 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 IPTC Web site, or in other forms from the IPTC.

Each list in the Subject Reference System will be maintained in an XML file that conforms to the definition of the TopicSet element in the NewsML DTD. 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">
        <TopicType Scheme="IptcTopicType" FormalName="Subject"/>
        <FormalName Scheme="IptcSubjectCodes">01000000</FormalName>
        <Description xml:lang="en" Variant="Name">Arts, Culture &amp;
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 item 8 below, 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. Additions to Subject Matter and other lists 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 a current IPTC member to sponsor the request.

Management of New Entries in the SRS

The addition of entries in the SRS may cause problems for both originators and receivers unless the correct approach is adopted.

Originators using these additions should, where appropriate, either advise the editorial staff of the relevant new entries and/or update language analysis software (when used) to include the new entry information.

Recipients must ensure their decoding system defaults to the appropriate higher level entry if they have not implemented the decode for new entries. This means that if an unrecognised SubjectMatter is received then the system should return the parent Subject and if an unrecognised SubjectDetail is received the system should return the parent SubjectMatter.

Changes to the Subject Reference System

Criteria for inclusion:

1: An IPTC member must need to use the term/s and gain support from other members for the inclusion during the process below.
2: The term should relate to general news, not to a specific discipline and have a universal meaning. The exception is the term that has come into global usage although its origin may be specific to a local discipline.
3: The term is unique in its definition and not a synonym of an existing term unless it falls under the definition of "Synonym" on page 7
4: No more than 20 new terms can be included in the "Fast Track" change process at any one time by any member.
5: Each new term must be accompanied by a precise explanation, in British English within the intended context of its use that will be included in the SRS
6: The Fast Track process can be used for both Subject Qualifiers and Subject Details.
7: All requests from non-IPTC members must be sponsored by an IPTC member in good standing.
8: Each new term requested should be in lower case, in the singular unless it is a plural noun and in British English.
9: Requests shall be made using the form found at
http://www.iptc.org/IPTC/metadata/documentation/IPTC-SRSsubmission.xls

Change Request Procedures Fast-Track (Subject Detail and Subject Qualifier)

1: Change requests to be sent to the Managing Director by e-mail (mdirector@iptc.org).
2: No more than 20 new terms will be accepted per request.
3: Accelerated process (similar to ISO fast track) maybe implemented for Subject Detail and Subject Qualifiers only. Fast track takes 28 days.
4: A jury of at least three but no more than five will consider change requests for Subject Detail and Subject Qualifiers only. The chairman of the Standards Committee will appoint the jury with the concurrence of the committee itself (These members should have a good knowledge of the SRS and its background)
5: A Change Request will be circulated to all members by the Managing Director. Members have 21 days to forward comments and constructive objections to the Managing Director from the request being circulated. The Managing Director will forward any comments to the jury at the end of the comment period.
6: Within 7 days following the 21 day "Comment" period the jury will either:
   * Agree the request as is.
   * Agree the request with minor modifications and agree with the requestor that their needs are still met.
   * Refuse the request but provide suggestions for a re-submission that could be agreed, or provide reasons why the proposal is not acceptable considering the criteria above. A resubmission will need to go through the full process again.
7: The jury spokesperson will inform the Director and the News Metadata Working Party Chair of its decisions at the end of the 7 day period.
8: If agreed by the jury, the Managing Director will assign a SubjectCode to each new term and update the SRS and its version number and post the updated version on the IPTC web site as soon as possible after the jury's decision.
9: A change request refused by the jury under the fast track process may be re-submitted for consideration under the non-fast track procedure below.

Change Request Procedures Standard-Track (Subject Matter, Subject Detail and Subject Qualifier)

1: Change requests to be sent to the Managing Director by e-mail (mdirector@iptc.org).
2: The Change Request is circulated to the membership inviting comments to be sent to the Director within a timeframe specified by the Director but not less than 21 days.
3: An ad-hoc Working Party of at least three but no more than five will consider the requests, reviewing suggestions and comments with the requesting organization, possibly changing a Subject Matter request to a Subject Detail, re-circulating a modified version to membership for comments. If agreement cannot be reached between the ad-hoc WP, the requestor way submit the change request directly to the MNA WP for

debate at the next full WP meeting

4: Once agreement is reached between the ad-hoc WP and the requesting organisation, the final result is circulated to all members and the ad-hoc WP will request the NMA WP to approve a motion to the Standards Committee that the terms and explanations contained in this paper be included in the SRS. The Standards Committee will vote on the motion at their next meeting.

5: The Managing Director will assign SubjectCodes to each new term and update the SRS and its version number and post the updated version on the IPTC web site as soon as possible after the Standards Committee approval.

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 levels 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 Web 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 System 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.

Examples

MySchemeSubject TopicSet controlled vocabulary is published by a Provider in a NewsItem under its own NewsML urn:

urn:newsml:myscheme.org:20001212:MySchemeSubjects:1

The new entry would be published as shown below. An extract from the current IptcSubjectCode file is provided first for comparison.

```
<TopicSet>
   . . .

<Topic Duid="sr04015000">
      <TopicType FormalName="SubjectMatter"/>
      <FormalName Scheme="IptcSubjectCodes">04015000</FormalName>
      <Description xml:lang="en-GB" Variant="Name">Transport</Description>
      </Topic>
   <Topic Duid="sr04015001">
      <TopicType FormalName="SubjectDetail"/>
      <FormalName Scheme="IptcSubjectCodes">04015001</FormalName>
      <Description xml:lang="en-GB " Variant="Name">Air Transport</Description>
      </Topic>
   <Topic Duid="sr04015002">
      <TopicType FormalName="SubjectDetail"/>
      <FormalName Scheme="IptcSubjectCodes">04015002</FormalName>
      <Description xml:lang="en-GB " Variant="Name">Railway</Description>
      </Topic>
   <Topic Duid="sr04015003">

<TopicType FormalName="SubjectDetail"/>
      <FormalName Scheme="IptcSubjectCodes">04015003</FormalName>
      <Description xml:lang="en-GB " Variant="Name">Road Transport</Description>
      <Description xml:lang="en-GB " Variant="Explanation"/>
      </Topic>
   <Topic Duid="sr04015004">
      <TopicType FormalName="SubjectDetail"/>
      <FormalName Scheme="IptcSubjectCodes">04015004</FormalName>
      <Description xml:lang="en-GB " Variant="Name">Waterway &amp; Maritime
Transport</Description>
      </Topic>

<!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, the last six 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 are:

```
<!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
[TopicSet] vocabulary.
        -->

<!ELEMENT tobject.property EMPTY>
        <!ATTLIST tobject.property
                %global-attributes;
                tobject.property.type CDATA   "current"
                >
<!--
                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
[TopicSet] 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 [TopicSet] 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 extract below explains implementation of the Subject Reference system within IIM Record 2.

Dataset No.	Dataset Name	Description
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. 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 and 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.

IPR	ES	Subject Reference Number	ES	Subject Name	ES	Subject Matter Name	ES	Subject Detail Name
				Subject Reference (SR) Dataset Structure *Minimum of 13 and maximum of 236 Octets*				
1 - 32 Octets registered with IPTC for UNO	:	Eight octets assigned by the IPTC as contained in Appendices H, I and J	:	Maximum 64 octets to reference the Subject	:	Maximum 64 octets to reference the Subject Matter	:	Maximum 64 octets to reference the Subject Detail
Minimum 1, maximum of 32	*1*	*8*	*1*	*0-64 Octets*	*1*	*0-64 Octets*	*1*	*0-64 Octets*

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 e-mail list or a registered holder of the Information Interchange Model, and

      c.   The requested IPR mnemonic has not been registered by any other Information Provider.

Application should be made to the Managing Director IPTC, by e-mail or fax, indicating
the Information Provider's full name, the desired IPR and a text description of the Company, Publication or Service making the submission. Both the IPR string and the description must be unique for each entry. When accepted, the IPR details will be published in the NewsML Provider TopicSet (find more information at http://www.iptc.org/metadata)

Legacy Formats

For legacy formats, where no numeric coding for subjects is possible, a single term can be used since there are implicit relationships 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 could be either entertainment or Business.

Below are notes on the two primary legacy transmission formats for news content.

ANPA 1312

It appears unlikely that any part of the Subject Reference System 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). See www.naa.org.

IPTC 7901

This recommendation is designed to assist users who make use of the keyword line is 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 cannot be added as new item to the format. Instead, 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:
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.

[End of document]